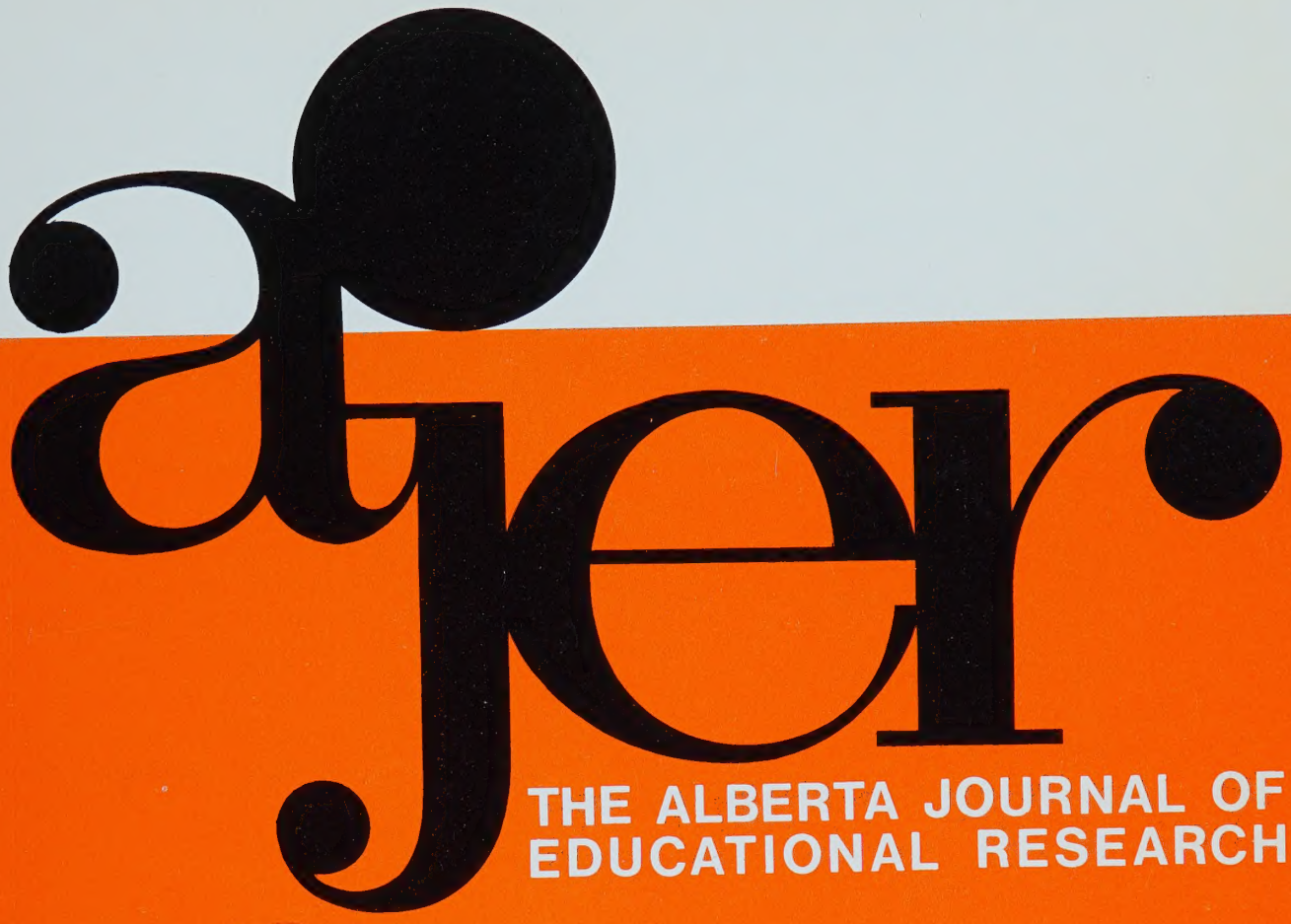

THE ALBERTA JOURNAL OF EDUCATIONAL RESEARCH

VOLUME XXVI NUMBER 1 MARCH 1980

PUBLISHED BY
THE UNIVERSITY OF ALBERTA • EDMONTON

THE ALBERTA JOURNAL OF EDUCATIONAL RESEARCH

A quarterly journal devoted to the dissemination, criticism, interpretation and encouragement of all forms of systematic enquiry into education and fields related to or associated with education.

Published quarterly in March, June, September, December by the
Faculty of Education, The University of Alberta

CONSULTING EDITORS

J. Britton
University of London

J. Calam
The University of British Columbia

M. Connelly
The Ontario Institute for Studies in Education

K. De Clerck
State University of Ghent

R. N. Evans
University of Illinois at Urbana-Champaign

R. H. Farquhar
University of Saskatchewan

E. Gagné
University of Ottawa

G. Harman
University of Melbourne, Australia

S. Hunka
The University of Alberta

J. W. G. Ivany
Simon Fraser University

D. A. MacIver
University of New Brunswick

L. D. Nelson
The University of Alberta

W. C. Nesbit
Memorial University of Newfoundland

E. Pedersen
McGill University

EDITOR: H. W. Hodysh

SECRETARY: A. Onishenko

FACULTY PUBLICATIONS COMMITTEE

M. A. Assheton-Smith
T. P. Atkinson
N. C. Bhattacharya
R. C. Bryce
W. T. Fagan
H. W. Hodysh
R. G. Martin (Chairman)
E. Miklos
V. R. Nyberg
C. H. Preitz

Editorial policy and the discussion and disposition of manuscripts are the joint responsibility of the Publications Committee. The views expressed and the accuracy of the statements made are the responsibility of the individual authors. The editor is solely responsible for the editorial comments.

AJER gratefully acknowledges support from the Social Sciences and Humanities Research Council of Canada and from the Alberta Advisory Committee for Educational Studies.

AJER is indexed in the *Canadian Education Index, Current Contents/Social and Behavioral Sciences,* and *Social Science Citation Index;* appropriate articles are abstracted in *Educational Administration Abstracts, Psychological Abstracts, Sociology of Education Abstracts,* and *Language Behavior Abstracts.*

The subscription rate is $8.00 per year; single copies are $2.50 each. Please make cheques payable to *The Alberta Journal of Educational Research.* All back issues are available; rates supplied on request. Claims for undelivered copies must be received within three months of the month of publication.

Address all communications and manuscript submissions to the Editor, *The Alberta Journal of Educational Research*, Faculty of Education, 732 Education South, The University of Alberta, Edmonton, Canada, T6G 2G5.

SECOND CLASS MAIL REGISTRATION NUMBER 1436

ISSN 0002-4805

The Alberta Journal of Educational Research

Volume XXVI, Number 1 *March, 1980*

CONTENTS

Prediction of Success in Student Teaching as a Criterion for Selection in Teacher Education Programs 1
R. J. Twa and M. Greene

A Critical Look at the WISC-R with Native Indian Children 14
B. Seyfort, O. Spreen and V. Lahmer

Cognitive Interactions in Science: Classroom Practices and Some Prescriptions of Theories of Learning 25
R. G. Hacker

The Validity of the *Estes Reading Attitude Scale* for Intermediate Grades 36
E. G. Summers

Predicting the Educational Bias Effect 44
G. G. Foster, B. Algozzine and R. C. Eaves

PERSPECTIVES

Research Traditions in the History of Education: A Study of Goals and Their Emergence 51
H. W. Hodysh

ESSAY REVIEWS

Teaching Educational Statistics in the Computer Age 60
R. E. Traub

RESEARCH NOTES

Reliability of the Alberta Essay Scales 64
V. R. Nyberg and A. M. Nyberg

BOOK REVIEWS

The Reform of Teacher Education in the United Kingdom, J. Lynch 68
Reviewed by James Britton

Teaching Canada for the '80s, A. B. Hodgetts and P. Gallagher 70
Reviewed by D. Massey

Educational Psychology, D. Johnson 73
Reviewed by F. C. Rankine

FACULTY OF EDUCATION
The University of Alberta

Digitized by the Internet Archive
in 2024 with funding from
University of Alberta Library

https://archive.org/details/ajer_1980

The Alberta Journal of Educational Research *Vol. XXVI, No. 1, March, 1980*

R. JIM TWA
and
MYRNA GREENE
The University of Lethbridge

Prediction of Success in Student Teaching as a Criterion for Selection in Teacher Education Programs

If a faculty of education is to select the best candidates for its teacher education program, it is necessary to be able to predict performance in student teaching. To determine their usefulness in predicting student teaching performance, five biographical particulars and forty personality test scores, obtained prior to candidates' entry into the education program, were examined in this study. Although the nine variables which entered the prediction equation for the total group accounted for only 37 percent of the variance in student teaching grade, much more promising results were obtained when the group was subdivided by sex, and subject specialization. The results indicated that the use of any single selection criterion, such as GPA, should be discontinued in favour of multiple selection criteria combined mechanically to produce a prediction of the candidates' grades in student teaching.

Every profession attempts to select for its training program those candidates who will become the best professionals in their particular field. Historically, teacher education institutions have had the reputation of being little concerned about who should enter the teaching profession. There is no intention to imply that faculties of education do not have entrance requirements. Most of the institutions do have selection policies, but very few limit enrollments. Haberman (1972) in his survey found that most of the 386 NCATE (National Council for Accreditation of Teacher Education) member institutions in the United States had some sort of

---

Mrs. Greene has been employed for the past three years in the Faculty of Education Research Centre at the University of Lethbridge. She received a Bachelor of Education Degree from the University of Lethbridge in 1976. Prior to that, she spent several years as an instructor in medical-surgical nursing.

Jim Twa is a Professor of Education at the University of Lethbridge where he has instructed in clinical supervision, analysis of teaching and practicum aspects of the teacher education program for the past nine years. He is presently coordinator of the University of Lethbridge's early field experience practicum which serves as one of the screening processes for students entering the Faculty of Education.

entrance requirements. Three hundred and forty-four of the institutions which constituted his sample relied heavily upon a minumum cumulative grade point average (GPA) and admitted almost everyone who had attained the minimum. If the alleged inflation of college grades ("Too Many A's," 1974) actually has occurred over the past few years, it follows that there are increased numbers of eligible, and generally lower quality, candidates being educated in ever-expanding programs. This condition, not conducive to producing excellent graduates, was discussed by Wisniewski at the 1978 Association of Teacher Educators' Annual Conference (Wisniewski, 1978).

In a recent comprehensive review of what the author (Menges, 1975) terms "assessing readiness for professional practice," several generalizations are made:

> First, definitions of effective practice should emphasize many discrete behaviors and characteristics rather than global definitions. Second, measures of these characteristics (predictors) should be as similar to the criterion itself as possible. Third, multiple assessment devices should be used so that no single type is overemphasized. Measures of subject matter knowledge, in particular, should be less heavily weighed. Fourth, data should not be used for decision-making until longitudinal studies demonstrate adequate predictive validity. (p. 201)

Teacher educators are not ignorant of the problems associated with present selection policies, and there appears to be increasing dissatisfaction with the situation. Two of the major speakers, Bruce Joyce and Karl Openshaw, at the 1978 Association of Teacher Educators' Annual Conference asserted that for significant improvement to occur in the education of teachers there must be a move toward improved selection procedures. In the face of the surplus of teachers, limited budgets, and the prospect of the necessity to limit enrollments, it would seem obvious that the institutions responsible for the education of teachers must increase their effort to develop selection criteria which are related to success in the field. The ability of a faculty to predict the success of its candidates from data available prior to their entry into the faculty would be a tremendous asset. The institutions' resources could then be invested in the most promising candidates, and students who were identified as having low potential for success in teaching would not waste their time and money pursuing an ill-chosen career.

With the present trend toward increased emphasis on the field experiences portion of teacher education programs, performance in student teaching is likely to become, if it is not already, the most important criterion in certification decisions. Faced with the oversupply of teachers and the consequent prospect of limited enrollments, the question of which students to admit becomes increasingly perplexing to schools of education. Prediction of performance in student teaching has been investigated by a considerable number of researchers employing a wide range of predictor variables. However, many of the studies dealt with only one category of predictor variable, such as grades, or personality factors, and many were one-shot research projects which were not replicated, thereby violating most of Menges' generalizations.

College grades, although the most widely used criterion for selection in teacher education programs, has received surprisingly little research attention as a predictor of student teaching performance. Those studies that have examined this predictor in various forms have generally reported a significant positive correlation between it and performance in student teaching (Greene, 1977; Emanuel, Laramore

& Sagan, 1975; Williams & Fox, 1967), but when used by itself, it has not proven to be particularly effective.

Another widely used criterion for selection which has received very limited attention is English proficiency. Bedford (1971) found that a measure of written English had some predictive value when it was combined with other variables. Two others (Cortis, 1968; Warburton, Butcher & Forrest, 1963) using a somewhat related predictor, the vocabulary portion of intelligence tests, found it of no value. Intelligence tests in general were found to be poor predictors in these same two studies.

Burroughs (1958), after making a rather detailed study of the interview, concluded that it was as good as any other predictor, a finding which was duplicated by Bedford (1971). Attitudes toward education, as measured by the Minnesota Teacher Attitude Inventory (Cook, Leeds & Callis, 1951), have been examined for their predictive value in a number of studies (Tarpey, 1965; Michaelis, 1954; Bedford, 1971; Williams & Fox, 1967). The first three studies reported the MTAI to have some predictive value while the fourth found it of no value. Tarpey and Bedford also found the Kuder Preference Record (Form C, 1960) to have very little predictive power.

In the area of personality and psychological measures, the Sixteen Personality Factor (Cattel, Elier & Tatsuoka, 1970) has been researched extensively (Davis & Satterly, 1969; Henjam, 1969; McClain, 1968; Morris, 1975; Tarpey, 1965; Warburton et al., 1963) and although the findings are not in complete agreement, all of the studies reported that it was one of the best predictors of success in student teaching. Michaelis (1954), using the Minnesota Multiphasic Personality Inventory, and Tyler (1954) using the MMPI and the Heston Personal Adjustment Inventory found them to be of little value as predictors. However, Gough, Durflinger and Hill (1968) reported considerable success in developing a prediction equation based on the scales of the California Psychological Inventory (Gough, 1957). Two other researchers (Veldman & Kelley, 1965) also using the CPI, identified significant test differences between student teachers who were rated high and those rated low in student teaching performance.

There appear to be conflicting results from the research in this area. However, by comparing some of the earlier studies (Sandiford, Cameron, Conway & Long, 1937; Michaelis, 1954; Tyler, 1954) to some of the more recent (Gough et al., 1968; McClain, 1968; Morris, 1975), it is possible to discern the progress which has been made. Later studies attempted to predict performance in student teaching based on different personality configurations and characteristics for male and female subjects teaching in different situations, for example, elementary versus secondary level. The results indicate that these types of studies are producing significant findings and should be continued.

The present study was based on two critical assumptions: (1) that it is not necessary to solve completely the "teacher effectiveness" problem before significant research can be done on the selection of candidates, and (2) that performance in student teaching is, or should be, the most important criterion upon which certification decisions are based.

Faced with the realities of selecting and counseling teacher education candidates, the investigator conducted a study to determine the feasibility of predicting performance in student teaching, based on data collected prior to the students' admission to the teacher education program.

Method

The study is a part of an ongoing, longitudinal project of the type recommended by Menges, currently underway at the University of Lethbridge (Dravland & Greene, 1978). That larger study provided for the creation of a data bank which contains up to 800 variables on each of about 700 education students in three populations, created for purposes of replication and validation. The data bank contains information on the students' biographical particulars, their two years in arts and science courses (students at the University of Lethbridge are required to spend two years in the Faculty of Arts and Science before being admitted to the Faculty of Education), scores on various English and psychological tests, performance in courses and practica in education, and data on their performance as teaching graduates. The data for this prediction study were drawn from the second population of that data bank. Data from a third population will be used for the purpose of validating the prediction equations developed in this study.

The Sample

The subjects for this study were the 180 students who were enrolled in the first course in the teacher education program at the University of Lethbridge, Lethbridge, Alberta, Canada, in the four semesters beginning Spring Semester, 1974 through to the end of Summer Session 1, 1975. During the first week of the course, the tests were administered to those students in three two-hour testing sessions. The biographical particulars were extracted from students' records. As these students progressed through the regular teacher education program, their performance in student teaching was evaluated by the university supervisors who observed the classroom performance of the students for a minumum of two hours per week during their six-week full-time student teaching experience. By the time the data had been assembled, the number of students for whom there were no missing data had decreased to 141 — 74 females and 67 males. Of the 141 subjects, 69 were classified as elementary and 70 as secondary, with two students unclassified as to route of preparation. This resulted in a further decrease to 139 subjects for any analysis involving these two subgroups.

The Variables

The variables chosen for this study included five biographical particulars, 40 test scores, and one measure of student teaching effectiveness:

1. Biographical Particulars:
   - Sex
   - Route — Elementary vs. Secondary Preparation
   - Age
   - Admission GPA — Cumulative GPA at time of entry into the first course in Education
   - Major GPA — the GPA in the teaching major at the time of entry into the Faculty of Education
2. Test Scores:
   - MTAI — percentile rank
   - D-Scale — total score on a Dogmatism scale
   - Q-Sort — score on an adjective Q-sort designed to measure self-concept
   - 16PF — raw scores on each sub-factor
     - A 16PF — Cyclothymia vs Schizothymia Factor
     - B 16PF — Intelligence Factor

C 16PF — Ego Strength Factor
E 16PF — Dominance Factor
F 16PF — Surgency Factor
G 16PF — Conscientiousness Factor
H 16PF — Extraversion Factor
I 16PF — Sensitivity Factor
L 16PF — Paranoid Factor
M 16PF — Unconventionality Factor
N 16PF — Sophistication Factor
O 16PF — Guilt Proneness Factor
Q_1 16PF — Radicalism Factor
Q_2 16PF — Self-sufficiency Factor
Q_3 16PF — Self-control Factor
Q_4 16PF — Id Pressure (Tenseness) Factor

TAT — achievement motivation score from the Thematic Apperception Test
College English — Placement Test — percentile rank
Convergent Question — score on a Convergent Question Test
CPI — standard score on 18 sub-scales.

Do CPI — Domination Scale
Cs CPI — Capacity for Status Scale
Sy CPI — Sociability Scale
Sp CPI — Social Presence Scale
Sa CPI — Self-Acceptance Scale
Wb CPI — Sense of Well-Being Scale
Re CPI — Responsibility Scale
So CPI — Socialization Scale
Sc CPI — Self-Control Scale
To CPI — Tolerance Scale
Gi CPI — Good Impression Scale
Cm CPI — Communality Scale
Ac CPI — Achievement via Conformance Scale
Ai CPI — Achievement via Independence Scale
Ie CPI — Intellectual Efficiency Scale
Py CPI — Psychological Mindedness Scale
Fx CPI — Flexibility Scale
Fe CPI — Femininity Scale

3. Grade in Student Teaching[1] — final grade in a six-week student teaching practicum

Numerical Value	Letter Grade
1	F, I, W (Failing, Incomplete, Withdrew)
3	C (Average)
4	B (Superior)
5	A (Excellent)

Limitations

A stepwise multiple regression procedure was used to "isolate a subset of available predictor variables that will yield an optimal prediction equation with as few terms as possible" (Nie, Hull, Jenkins, Steinbrenner & Bent, 1975, p. 345). In the initial analysis emphasis was on producing the most effective "mechanical" model for the selection of teacher candidates. There are obvious ethical and measurement problems associated with selection schemes and quotas. There was no attempt in this study to deal with these issues. Nor was there any attempt to analyze the psychological characteristics implicit in the regression equations. The practicality of applying the model will be dealt with in a further study.

For a variable to be included in the regression analysis, the F value for the difference between multiple correlations had to be significant at the .05 level. The degrees of freedom used to test the significance were $(m_1 - m_2)$ and $(N - m_1 - 1)$ where m_1 is the larger number of independent variables and m_2 the smaller number of independent variables (Guilford & Fruchter, 1973, p. 368).

Also included in the discussion are those variables which are correlated with the criterion at the .05 level of significance. The sample was examined first as a total group and was then subdivided for further analyses.

Results

The Total Group

The means and standard deviations of the criterion and predictor variables for the total group are presented in Table 1. Space considerations do not permit inclusion of the complete correlation matrix; however, it is available to any researcher wishing to reproduce this work.

Of the total 45 predictor variables available, only 12 qualified for consideration (See Table 2). With the F value for inclusion set at 3.92, the regression equation produced a multiple correlation of .609 with a standard error of .82254. The predicted grade in student teaching was as follows:

$$\begin{aligned}\text{Grade} = {} & 1.64481 + (-.00255)\text{FxCPI} + \\ & (.16554)\text{Q}_1\ 16\text{PF} + (.54311)\text{Major GPA} + \\ & (.10243)\text{F16PF} + (.02184)\text{DoCPI} + (-.03121)\text{AiCPI} \\ & + (.02964)\text{CmCPI} + (-.11572)016\text{PF} + (-.02129)\text{SyCPI}\end{aligned} \tag{1}$$

Examining the makeup of the regression equation and the order in which the predictor variables entered the equation, one might wonder why a variable such as the Q-Sort which was correlated at the .01 level of significance with Grade in Student Teaching, did not enter the equation, while So CPI with a correlation of .092, did enter. There are two conditions under which the variables may prove to be valuable as predictors in the multiple regression process. The first occurs when two predictor variables have high correlations with the criterion and low correlations with one another. The other situation in which a variable may prove to be valuable in the equation occurs when it acts as a suppressor on one of the other predictor variables which has previously entered the equation. In this case the suppressor variable has a low correlation with the criterion and a high correlation with one of the other predictor variables.[2] Examples of these situations are shown in Table 3 which identifies the intercorrelations of the variables in question. For example, the Q-Sort was significantly correlated ($r = -.206$) with Fx CPI which entered the equation on Step 1. Apparently Q-Sort was accounting for much of the same variance in the criterion as was Fx CPI and therefore had its predictive value lowered when Fx CPI entered the equation. On the other hand, Sy CPI which produced a low correlation with Grade in Student Teaching ($r = .092$) did enter the equation. Sy CPI was correlated with one of the variables which had previously entered the equation, Cm CPI ($r = .118$), and therefore appeared to be acting as a suppressor on that variable. In this way it became a useful predictor in the equation while the Q-Sort did not.

Dividing the Sample by Sex and Route

Two previous studies (Gough et al., 1968; McClain, 1968) found indications that there are different personality configurations for successful male and female

TABLE 1
MEANS AND STANDARD DEVIATIONS OF CRITERION AND PREDICTOR VARIABLES FOR THE TOTAL SAMPLE ($N = 141$)

	Mean	Standard Deviation
Criterion		
Grade in Student Teaching	3.773	1.003
Biographical Predictors		
Sex	1.475	0.501
Age	23.638	4.654
Admission Grade Point Average	2.711	0.473
Major Grade Point Average	2.829	0.516
Route	1.504	0.502
Psychological Test Score Predictors		
MTAI	47.220	28.700
D Scale	101.972	25.291
Q Sort	66.674	19.895
A 16PF	5.355	1.559
B 16PF	6.234	1.442
C 16PF	5.986	1.732
E 16PF	5.064	1.810
F 16PF	5.348	1.801
G 16PF	5.851	1.748
H 16PF	5.688	1.860
I 16PF	5.582	1.778
L 16PF	5.312	2.022
M 16PF	6.759	2.032
N 16PF	5.411	1.882
O 16PF	5.539	1.973
Q_1 16PF	5.901	1.947
Q_2 16PF	5.376	2.212
Q_3 16PF	6.433	1.928
Q_4 16PF	4.957	1.736
TAT	11.518	6.211
Col Eng	60.355	25.237
Con Quest	2.730	1.303
Do CPI	53.589	10.970
Cs CPI	51.248	10.146
Sy CPI	51.865	9.540
Sp CPI	56.596	10.268
Sa CPI	56.936	10.287
Wb CPI	49.213	9.081
Re CPI	45.652	9.179
So CPI	48.716	10.088
Sc CPI	46.844	10.202
To CPI	50.270	9.338
Gi CPI	46.801	10.384
Cm CPI	51.262	7.448
Ac CPI	50.723	9.820
Ai CPI	55.589	8.731
Ie CPI	51.823	9.367
Py CPI	53.355	8.848
Fx CPI	55.844	11.071
Fe CPI	51.504	10.083

TABLE 2

CORRELATIONS BETWEEN PREDICTORS AND GRADE IN STUDENT TEACHING, AND RESULTS OF MULTIPLE REGRESSION ANALYSIS (N = 141)

Predictors	Correlation with Grade in Student Teaching	Order of Entry into Regression Equation	Resulting R
Biographical Particulars			
Sex (Male)	-.182*		
Admission GPA	.216		
Major GPA	.255**	(3)	.432
Psychological Test Scores			
Q-Sort	.243**		
F 16PF	.111	(4)	.477
O 16PF	-.216*	(8)	.591
Q_1 16PF	.267**	(2)	.377
Do CPI	.243**	(5)	.513
Sy CPI	.092	(9)	.609
Cm CPI	.205*	(7)	.566
Ai CPI	-.189*	(6)	.542
Fx CPI	-.275**	(1)	.275

* $p < .05$
** $p < .01$

TABLE 3

INTERCORRELATION MATRIX (N = 125)

Variable Names	Grade in Student Teaching	Fx CPI	Ai CPI	Cm CPI	Sy CPI	Do CPI	Q_1 16PF	O 16PF	F 16PF	Q-Sort	Major GPA	Adm GPA
Sex (Male)	-.182	.156	.017	-.028	.099	.101	-.120	-.160	-.113	.181	-.147	-.168
Adm GPA	.216	-.055	.206	.091	.050	-.020	.130	-.003	-.195	.102	.895	1.00
Major GPA	.255	-.049	.186	-.000	.066	.028	.133	-.053	-.127	.074	1.00	
Q-Sort	.243	-.206	.081	.264	.237	.279	.185	-.357	-.093	1.00		
F 16PF	.111	.006	-.343	.020	.295	.050	-.243	.039	1.00			
O 16PF	-.216	.052	-.152	.081	-.369	-.342	-.098	1.00				
Q_1 16PF	.267	-.028	.211	-.144	-.056	.031	1.00					
Do CPI	.243	-.112	.100	.207	.606	1.00						
Sy CPI	.092	-.028	.050	.118	1.00							
Cm CPI	.205	-.279	-.047	1.00								
Ai CPI	-.189	.438	1.00									
Fx CPI	-.275	1.00										

Note. For N = 125: .05 level of significance = .174
.01 level of significance = .228

student teachers, but Morris (1975) was unable to verify this finding in his study. Morris also was unable to identify differences between good elementary and secondary teachers and concluded that "the differences between good and poor teacher personality is the same regardless of sex or grade level taught" (p. 20). With a view to examining the discrepancies found in these previous studies, the investigator made additional analyses by dividing the sample by sex, by grade level taught, and by subject major or level or preparation. The following is a summary of a more complete analysis, the results of which are available from the author upon request.

Females Only (N = 74). Employing 44 variables (Sex excluded) as available predictors and applying the same criteria for consideration as in the previous analysis, a total of ten variables qualified for consideration in this section. Again, with the *F* value for inclusion set at 4.00, the regression equation produced a multiple correlation of .699 with a standard error of .76964. Six predictors entered the equation. The equation for predicting grade in student teaching was:

$$\begin{aligned}\text{Grade} = {} & 1.95426 + (.19797)Q_1\ 16PF + (.33980)\text{Major GPA} \\ & + (.04419)\text{CmCPI} + (-.04455)\text{AiCPI} + (-.01184) \\ & \text{D Scale} + (.02654)\text{ReCPI}\end{aligned} \tag{2}$$

There were several differences between this equation and Equation 1 for the total group. First, only four of the variables were common to both equations: Q_1 16PF, Major GPA, Ai CPI, and Cm CPI. Notable by its absence from Equation 2 was Fx CPI which entered on Step 1 of Equation 1. Also missing were F 16PF, Do CPI, O 16PF, and Sy CPI. Two variables, Q-Sort and Re CPI, entered Equation 2, but did not appear in Equation 1 for total group.

Males Only (N = 67). Using the same predictors and criteria, none of the biographical particulars and only five of the test scores qualified for analysis. The equation produced a multiple correlation of .511 and a standard error of .86165 with only two variables entering:

$$\begin{aligned}\text{Grade} = {} & 4.36615 + (.02105)\text{Q-Sort} \\ & + (-.0419)\text{PyCPI}\end{aligned} \tag{3}$$

Neither of these two variables appeared in the equations for the total group for females. Both were significantly correlated with the criterion ($p < .05$) but not with each other.

Elementary Only (N = 69). When the regression equations for the elementary-secondary groups of subjects were examined, an interesting but perplexing finding appeared. Sex entered on Step 2 of the equation for predicting grade in student teaching for secondary level subjects. Because of the practical nature of this study and the obvious implications which would arise from the use of sex as a criterion for the selection of candidates for teacher education programs, sex was removed from the array of available predictors and the data were reprocessed with the remaining 44 available predictors.

Five variables qualified for consideration in the analysis of elementary-route students, but only four entered the regression equation producing a multiple correlation of .685 and a standard error of .77172.

$$\begin{aligned}\text{Grade} = {} & -1.28909 + (.27778)Q_1\ 16PF \\ & + (.06249)\text{CmCPI} + (.48064) \\ & \text{Major GPA} + (-.02439)\text{IeCPI}\end{aligned} \tag{4}$$

Secondary Only ($N = 70$). With the same variables and criteria as for elementary- route students, three variables entered the regression equation, producing a multiple correlation of .528 and a standard error of .85689.

$$\begin{aligned}\text{Grade} = {} & 3.66568 + (.01504)\text{Q-Sort} \\ & + (-.04999)\text{AiCPI} \\ & + (.03515)\text{IeCPI}\end{aligned} \tag{5}$$

Dividing the Sample Into More Specific Combinations

Sex and Route. A further analysis was completed for each of female-elementary, female-secondary, male-elementary, and male-secondary students. Numbers in each group were relatively small; consequently it was possible to account for a large portion of the variance in student teaching grade. The multiple correlations and prediction equations produced were as follows:

Female-elementary — $N = 49$, $R = .768$, standard error = .73181

$$\begin{aligned}\text{Grade} \quad & -.53328 + (.29696)\text{Q}_1\ 16\text{PF} \\ & + (.07683)\text{CmCPI} \\ & + (-.00955)\text{D-Scale} \\ & + (-.03392)\text{TAT}\end{aligned} \tag{6}$$

Female-secondary — $N = 24$, $R = .698$, standard error = .54042

$$\begin{aligned}\text{Grade} = {} & 28264 + (-.06733)\text{SaCPI} \\ & (.99046)\text{Major GPA}\end{aligned} \tag{7}$$

Male-elementary — $N = 20$, $R = .596$, standard error = .68224

$$\text{Grade} = 6.37080 + (-.05282)\text{PyCPI} \tag{8}$$

Male-secondary — $N = 46$, $R = .570$, standard error = .87867

$$\begin{aligned}\text{Grade} = {} & 3.55811 + (.02643)\text{Q-Sort} \\ & + (-.03338)\text{PyCPI}\end{aligned} \tag{9}$$

Subject Major Subgroups. Pursuing one step further the hypothesis "that successful teaching in different fields calls for a personality unique to each field" (McClain, 1968, p. 31), three subject major subgroups — physical education, social studies, and English — were selected for analyses. These three subject areas had the largest numbers of students and had generally been oversubscribed in terms of available teaching personnel at the University of Lethbridge. Again, numbers were small and a large portion of the variance was accounted for. The regression equations were as follows:

Physical Education — $N = 24$, $R = .750$, standard error = .72174

$$\begin{aligned}\text{Grade} = {} & 4.69066 + (1.38098)\text{Major GPA} \\ & + (-.06163)\text{PyCPI} \\ & + (-.26474)\text{B16PF}\end{aligned} \tag{10}$$

Social Studies — $N = 32$, $R = .932$, standard error = .40064

$$\begin{aligned}\text{Grade} = {} & 3.04686 + (.05415)\text{Q-Sort} \\ & + (-.03806)\text{GiCPI} \\ & + (-.05446)\text{SaCPI} \\ & + (4.52233)\text{Admission GPA} \\ & + (-3.61880)\text{Major GPA} \\ & + (.14192)\text{B16PF} \\ & + (-.04004)\text{AcCPI} \\ & + (.12128)\text{C16PF}\end{aligned} \tag{11}$$

English — $N = 25, R = .900$, standard error = .46941

$$\begin{aligned} \text{Grade} = {} & -1.04184 + (.05282)\text{DoCPI} \\ & + (.18464)\text{Q}_1\ 16\text{PF} \\ & + (.15745)\text{L16PF} \\ & + (-.2567)\text{E16PF} \\ & + (.23055)\text{F16PF} \end{aligned} \tag{12}$$

Discussion

The major purpose of this study was to determine which of the variables, if any, were worthy of further examination as predictors of performance in student teaching. In the next phase of this study the prediction equations will be cross-validated on another population of students by comparing their predicted grade in student teaching with their actual grade.

From the findings of this study, it was apparent that no single variable by itself was useful as a predictor. The best single predictor for the total group was Fx CPI, which accounted for only 7% of the variance in student teaching grade. Even for certain subgroups where there was a fairly high correlation between the criterion and one predictor (e.g., DoCPI at .641 for English majors), the proportion of variance accounted for was not high enough to be useful.

Other interesting findings were certain "non-relationships." For example, neither the College English Placement Test score nor the MTAI were of any value as predictors of grade in student teaching for any of the groups in this study. Several of the subscales on the 16PF and the CPI, which from their descriptions were expected to be useful predictors, did not enter any of the regressions. In other instances, two subscales with apparently conflicting descriptions were both significantly correlated with the criterion variable. These discrepancies can be accounted for by examining differences among groups.

Another misleading finding which existed for the total group was a "masking" effect. An example of this was in the D-Scale (dogmatism). The correlation with the criterion for the total group was –.028, leading one to believe that dogmatism was randomly distributed with respect to grade in student teaching. However, for the female-elementary subgroup the correlation was –.340 ($p < .05$), and for the male-elementary subgroup the correlation was .535 ($p < .05$), thus producing a masking effect for the total group. A similar effect existed for other variables such as Major GPA.

The findings of this study were to be regarded only as hypotheses to be tested in the next stage of the project. Also, for those subgroups with a very small sample size, the procedures employed may have capitalized on artifacts of the data. Keeping these concerns in mind there appear to be two general findings worthy of note:

1. The more specific one becomes in specifying the sex, grade level taught, and subject taught, the more accurate the prediction of grade in student teaching.
2. To become more accurate in predicting grade in student teaching, one must include as predictors psychological and personality test results in addition to biographical data.

A desirable feature of any selection procedure is that it employ the minimum number of predictors which would yield the highest possible prediction. The decision to include or exclude a certain variable in the practical application of a

prediction model necessitates weighing the increase (or decrease) in predictive power against the expenditures required to obtain the data for that variable. One must also consider the ethical and moral implications of the "invasion of privacy issue" raised when any tests are used for selection purposes. No attempt to deal with such issues has been undertaken in this paper.[3] All of the regression equations reported in this study must be cross-validated on a subsequent group of subjects before more than tentative confidence can be placed in the results. However, the findings of this study indicate that further investigation in this area is definitely warranted.

Notes

1. To date there is no measure of the reliability or validity of the grade in student teaching. Other studies within the longitudinal project are examining those concerns.
2. For expansion of this idea, see Wiggins (1973), Chapter 1, "The Model and Its Application."
3. The authors have created several hypothetical prediction models based on a rational inclusion or exclusion of various variables.

The authors wish to acknowledge the personnel of the Faculty of Education Research Centre at the University of Lethbridge and the Research and Development Centre for Teacher Education, University of Texas at Austin for their assistance with the preparation of this report.

References

Bedford, C. M. Practice teaching: Ten years of screening out weak prospective teachers. *Saskatchewan Journal of Educational Research and Development,* 1971, *2* (1).

Burroughs, G. E. R. A study of the interview in the selection of students for teacher training. *British Journal of Educational Psychology,* 1958, *28,* 37-46.

Cattel, R. B., Elier, H. W., & Tatsuoka, M. M. *Handbook for the sixteen personality factor questionnaire.* Champaign, Ill.: Institute for personality and ability testing, 1970.

Cook, W. W., Leeds, C. H., & Callis, R. *The Minnesota teacher attitude inventory.* New York: Psychological Corporation, 1951.

Cortis, G. A. Predicting student performance in colleges of education. *British Journal of Educational Psychology,* 1968, *32(2),* 115-122.

Davis, T. N., & Satterly, D. J. Personality profiles of student teachers. *British Journal of Education Psychology,* 1969, *39,* 183-187.

Dravland, V., & Greene, M. *Development of a model for the evaluation of teacher education programs* (QAULTEP). Paper presented at the American Educational Research Association annual meeting, San Francisco, April, 1979.

Emanuel, J., Laramore, D., & Sagan, E. The relation of selected academic program variables to student teaching performance. *Contemporary Education,* 1975, *46* (4), 245-248.

Gough, H. G. *Manual for the California psychology inventory* (Rev. ed.), Palo Alto, CA: Consulting Psychologists Press, 1957.

Gough, H. G., Durflinger, G. W., & Hill, R. E. Predicting performance in student teaching from the California psychological inventory. *Journal of Educational Psychology,* 1968, *59,* (2), 119-127.

Greene, M. L. *A study of relationships between criteria used to predict potential success in student teaching and criteria used to assess actual success in student teaching.* Unpublished manuscript. University of Lethbridge, 1977.

Guilford, J. P., & Fruchter, B. *Fundamental statistics in psychology and education* (5th ed.). New York: McGraw-Hill Book Co., 1973, 368.

Haberman, M. *Guidelines for the selection of students into programs of teacher education.* Paper presented at the Association of Teacher Educators. Chicago, February, 1972.

Henjam, A. A study of the significance of student teachers' personality characteristics. *Journal of Teacher Education,* 1969, *20,* 143-147.

Joyce, B. *The theory practice gap.* Paper presented at the Annual Meeting of the Association of Teacher Educators, Las Vegas, Nevada, February, 1978.

McClain, E. W. Sixteen, p. f. scores and success in student teaching. *Journal of Teacher Education,* 1968, *19,* 25-32.

Menges, R. J. Assessing readiness for professional practice. *Review of Educational Reseach,* 1975, *45* (2), 173-207.

Michaelis, J. U. The prediction of success in student teaching from personality and attitude inventories. *University of California Publications in Education,* 1954, *11* (6), 415-484.

Morris, J. D. Personality and student teaching success. *Journal of Experimental Education,* 1975, *43* (4), 15-20.

Nie, N. H., Hull, C. H., Jenkins, J. G., Steinbrenner, K., & Bent, D. H. *Statistical packages for the social sciences* (2nd ed.). New York: McGraw-Hill, 1975.

Openshaw, K. *Procedures for training the pre-service teacher: The instructional strategy gap.* Paper presented at the Annual Meeting of the Association of Teacher Educators, Las Vegas, Nevada, February, 1978.

Sandiford, P., Cameron, M. A., Conway, C. B., & Long, J. A. *Forecasting teaching ability.* (Bulletin No. 8). Toronto, Ontario: University of Toronto, Department of Educational Research, 1937.

Tarpey, M. A., Personality factors in teacher trainee selection. *British Journal of Educational Psychology,* 1965, *35,* 140-149.

Too Many A's. *Time,* November 11, 1974, 88.

Tyler, F. T. The prediction of student teaching success from personality inventories. *Public Education,* 1954, *2* (4), 233-314.

Veldman, D. J., & Kelly, F. J. Personality correlates of a composite criterion of teaching effectiveness. *Alberta Journal of Educational Research,* 1965, *11,* 102-107.

Warburton, F. W., Butcher, H. J., & Forrest, G. M. Predicting student performance in a university department of education. *British Journal of Educational Psychology,* 1963, *33,* 68-79.

Wiggins, J. S. *Personality and prediction: Principles of personality assessment.* Reading, Mass.: Addison Wesley Publishing Co., 1973.

Williams, J., & Fox, A. M. Prediction of performance in student teaching. *Educational and Psychological Measurement,* 1967, *27* (4), 1169-1170.

Wisniewski, R. *Closing the credibility gaps: Resolution and responsibility.* Address presented to the Annual Meeting of the Association of Teacher Educators, Las Vegas, Nevada, January 31 to February 3, 1978.

B. SEYFORT
Glendale Lodge Society
O. SPREEN
and
V. LAHMER
University of Victoria

A Critical Look at the WISC-R with Native Indian Children

The Wechsler Intelligence Scale for Children (WISC-R) was administered to 177 Native Indian children from three different bands in southwestern British Columbia. Item responses and difficulty gradients were evaluated and discussed for nine of the subtests. The raw score distributions on each subtest for ages six through twelve years were converted into scaled scores with a mean of ten and a standard deviation of three (comparable to Wechsler's scaled scores for the original test). These scaled scores are available on request from the first author. A large number of items, especially from verbal subtests, were found not to contribute significantly to the total test variance. The implications of these findings are discussed in relation to clinical use of the WISC-R for Native Indian children.

The clinician charged with the psychological assessment of ethnic or minority groups is frequently faced with the problem of employing intelligence tests which have been standardized on a representative sample of the general population. Although a particular ethnic group may be part of the standardization sample, the pattern or profile unique to that group is lost in the overall population variance. Thus, intra-group comparison, i.e., interpretation of the individual's score with ref-

---

Bernice Seyfort, Ph.D., is a graduate of the Neuropsychology programme, University of Victoria. She is presently Director of Psychology for the Glendale Lodge Society. Clinical and research interests include language development, learning disorders, the use of nonverbal communication among the mentally and physically handicapped, and the development of community college programmes for training direct care staff to work with the mentally handicapped.

Otfried Spreen, Ph.D., is Professor of Psychology at the University of Victoria and Director of the Neuropsychology Laboratory of the University, as well as consultant to the Royal Jubilee Hospital. His research interests are in the areas of clinical neuropsychology with special focus on language disorders, learning disabilities and mental deficiency.

Victor C. Lahmer, B.Sc., graduated from the University of Victoria in 1978. He has a special interest in studies of Canadian Natives and has worked as a social work aid in Hay River, North West Territories.

erence to his age and ethnic group is rendered impossible from the norms published for the population at large.

The reasons for differences between groups of different ethnic origin in performance on tests of general intelligence have been the topic of wide speculation but are not a major focus of the present study. A number of parameters affecting differential performance among cultural groups have been identified. They can be roughly divided into test taking attitudes such as motivation to succeed, test anxiety, cultural values, and test content such as differential familiarity with certain tasks and particularly varying experience with verbal material in general.

Differential test performance between majority population and certain ethnic groups has been partialled out or dealt with in several ways. First, the addition of a constant to an individual's obtained score can be used to adjust for systematic differences from published norms (Droege, Botterbusch & Hawk, 1969). Second, ethnic and socioeconomic variables can be treated as moderators (Gaddes, McKenzie & Barnsley, 1968; Guion, 1966). Individuals are categorized on the basis of race, socioeconomic level, or other relevant factors with test scores and their interrelationships investigated within each group. A third approach involves constructing culture-fair tests with the objective of minimizing differences between ethnic and socioeconomic groups. The movement to develop culture-fair tests began in the 1950's but did not resolve the clinician's dilemma. To begin with, culture-fair tests vary considerably in terms of cognitive functions they purport to measure and further, they usually omit tasks related to school learning which are critical in the assessment of children.

The described problems are particularly apparent in the assessment of indigenous persons. The performance of Native Indian children on most standardized intelligence tests indicates dull normal functioning characterized by average perceptual-motor skills and poor verbal skills (Goodenough, 1962; Peters, 1963; Vernon, 1966; MacArthur, 1968, 1969).

MacArthur (1969) gave group-tests to North West Territory Indian and Métis youths. Average or above average scores were demonstrated on the performance tests while low scores were attained on the verbal tests. Other investigators using a variety of tests with a number of Native Indian groups also found these children to be superior to their White counterparts on performance tasks and inferior on verbal tasks.

Cazden and John (1971) cite several examples of superiority on tasks such as visual imagery, skills in interpreting photographs, learning through imitation, and the Draw-A-Man test. On the other hand, Gaddes, McKenzie and Barnsley (1968) found no statistical support for the hypothesis that North American Indian children possess superior spatial imagery as measured by a Wechsler Intelligence Scale for Children (WISC) subtest and the Porteus Maze Test. Brophy and Aberle (1966) suggest that the Native Indian concept of time, their disposition to conform to nature, and their withdrawal in school are three sources of value conflict which limit verbal abilities. It is also suggested that children who are reared in a culture which provides less experience of literacy behaviour and less awareness of the act of speech can be expected to experience greater problems on the verbal portion of intelligence tests as well as greater cognitive confusion about reading and writing (Downing, Ollila & Oliver, 1975). The problem is compounded by the presence of varying degrees of bilingualism in Native Indian children.

A Canadian study by Taylor and Skanes (1976) compared Inuit and White

children, matched for grade placement and environmental circumstances, on verbal-educational tasks, spatial abilities, and inductive reasoning. After only one year at school, no difference in verbal-educational tasks or inductive reasoning could be demonstrated. This finding, which emphasized the importance of environment on the development of abilities, has also been supported by Wilson (1973). In an early study, Rohrer (1942) hypothesized that if Native Indians could be matched with White children on socioeconomic status, cultural opportunities and comparable schooling, a difference in I.Q. scores would not be demonstrable. This hypothesis was confirmed.

The Wechsler Intelligence Scale for Children-Revised (WISC-R) (Wechsler, 1974) is a commonly accepted tool for evaluating general intelligence. However, relatively little research is currently available relating differential test responses to cultural groups. Such information regarding Native Indian children would be of value for a number of reasons. As a rule, psychologists consider not only total test and subtest performance, but often make clinical inference based on particular item responses. In addition, psychologists are interested in a scale which can reveal individual differences within a specific group. For these purposes it is necessary to know something of the discriminatory power of the individual items as well as the order of difficulty of items when the scale is applied to Native Indian children.

The WISC-R manual cites correlational studies of the test with the Wechsler Preschool Primary Scale of Intelligence (WPPSI), the Wechsler Adult Intelligence Scale (WAIS), and the Stanford-Binet, Form L-M. The major difference between the original (1949) and revised form of the WISC is that the standardization sample for the latter includes non-Whites and is therefore more representative than the original test. Furthermore, the WISC-R has new administration and scoring criteria with an altered sequence of subtest presentation. Both forms of the test yield a full scale verbal and performance I.Q. with a mean of 100 and a standard deviation of 15. Only 10 subtests of the WISC and WISC-R are considered mandatory. The I.Q. is calculated on the basis of five verbal subtests (Information, Similarities, Arithmetic, Vocabulary, and Comprehension) and five performance subtests (Picture Completion, Picture Arrangement, Block Design, Object Assembly, and Coding).

Few studies have examined item responses of the WISC for a given population. Items inappropriate for Canadians were identified in studies by Spreen and Tryk (1970) and Vernon (1976; 1977). WISC protocols were collected on ninety disadvantaged children from varying cultural backgrounds by Vance and Gaynor (1976). The full scale I.Q.'s ranged from 80 to 95. Analyses indicated that while item difficulties compared rather well with those obtained with the standardization sample, there did appear a need for rearrangement of some subtest items particularly on the Vocabulary and Picture Completion subtest. It was also evident with this population that a large proportion of the items did not contribute to the total test variance.

The purpose of the present study was to gather information on the performance of Native Indian children on the WISC-R. It was felt that this information would facilitate more accurate and meaningful assessment of Native children by clinicians serving similar populations. Specifically, the study was designed to compile mean raw scores for Native Indian children on the WISC-R, to evaluate the item responses and difficulty gradient of the WISC-R from the sample of Native children tested, and to restandardize the sample to comply with Wechsler's test constructed for nine tests in the battery.

Method

The subject population included 177 Native Indian children (93 boys and 84 girls) ranging in age from 6 to 13 years, representing three separate bands in Southwestern British Columbia; Native children lived on reserves and attended regular schools allowing for their integration into the wider community. The first band was represented by 95 children, the second by 14, and the third by 68.

The data were gathered in the summers of 1975 and 1977 by two male examiners. The WISC-R was given to each subject in a quiet area of the school or band office. Rapport with the children was established and maintained by the examiners outside as well as inside the testing situation by their continued presence among the Native children. Both examiners were housed on or near the reserves with the cooperation of the Native people. The purpose of the study was explained to Band representatives and feedback was given after testing had been completed. In the administration and scoring of the WISC-R the manual was carefully followed.

Results and Analyses

Data Analysis

Single item response frequency and difficulty gradient were analyzed by coding items as 0 for incorrect and 1 for correct responses; partial or qualitative scores were not recorded separately. Items were considered to have been correctly answered if any credit could be allowed. While this procedure is likely to reduce individual differences in item responses it was followed to increase ordinal consistency of the test items as well as to simplify the analyses.

For the restandardization of subtests, the distribution of raw scores at each age level was computed and converted to a scale having a mean of 10 and a standard deviation of 3. Minor irregularities at the extreme ends of each scale were eliminated by smoothing.

Effects of Age and Group Membership

An Age x Group (Band) Analysis of variance (ANOVA) was carried out on full scale scores to examine the effects of these two factors. The five thirteen-year-old subjects were omitted from the analysis as this age level was not represented in two of the three groups. Significant main effects were demonstrated for both age (F [6, 151] = 3.42, $p < .001$) and group (F [2,151] = 3,86, $p < .001$). The Age x Group interaction was not found to be significant. The group effect may represent true differences between the three Native Bands, however the size of one group was quite small ($N = 14$). For the purpose of subsequent analyses the data from all three groups were collapsed to supply a larger sample size. Table 1 presents the means, standard deviation and univariate tests of significance for each subtest across seven age levels.

Subtest Item Validity and Difficulty

Tables 2 to 6 present the results of the item analyses on each of the verbal subtests. These have been tabulated in three columns as follows: (a) frequency of subjects having answered the items correctly, (b) the proportion of subjects answering the item correctly, and (c) the point biserial correlation coefficient (pb) between each item and the total weighted score (index of discrimination). Subtest

TABLE 1

F VALUES, MEAN SQUARES, AND *P* VALUES FOR AGE FACTOR ON WISC-R SUBTEST RAW SCORES

Variable	*F*(7, 168)	Mean SQ	*p* less than
Information	20.28	165.42	0.00
Picture Completion	19.61	278.27	0.00
Similarities	9.56	171.81	0.00
Picture Arrangement	14.09	1085.90	0.00
Arithmetic	24.25	129.12	0.00
Block Design	13.08	1231.08	0.00
Vocabulary	11.79	495.38	0.00
Object Assembly	13.10	266.26	0.00
Comprehension	7.30	252.02	0.00

	Age													
	6 (*N* = 21)		7 (*N* = 23)		8 (*N* = 30)		9 (*N* = 19)		10 (*N* = 31)		11 (*N* = 29)		12 (*N* = 20)	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Information	4.14	1.76	6.34	2.63	6.96	2.87	8.89	2.86	9.87	3.58	10.96	2.41	11.85	2.25
Picture Completion	7.76	4.03	11.26	4.99	14.56	3.29	14.73	3.36	16.19	3.91	18.13	2.48	18.35	4.23
Similarities	3.09	2.82	5.13	4.14	6.26	3.68	6.52	3.94	8.12	5.24	10.06	4.16	11.40	4.01
Picture Arrangements	6.80	7.06	12.39	9.78	16.56	9.33	21.42	9.20	25.16	8.37	23.41	10.03	26.10	6.80
Arithmetic	4.71	2.21	6.21	2.17	7.66	2.56	9.21	2.17	10.09	2.73	11.10	2.27	11.90	1.86
Block Design	8.38	4.98	8.82	5.79	14.50	9.80	21.10	10.17	24.83	13.36	23.37	10.34	26.25	8.08
Vocabulary	11.85	4.69	16.08	6.45	17.73	5.99	19.42	5.91	22.41	8.21	24.72	6.29	25.55	5.44
Object Assembly	10.80	5.44	16.13	4.71	17.96	4.90	20.42	2.36	20.61	4.95	21.55	4.18	21.40	4.29
Comprehension	5.28	2.51	8.21	3.28	9.53	3.87	11.21	3.89	14.74	11.29	13.68	3.86	14.55	4.14
Coding	23.10	11.07	31.00	13.98	22.88	8.01	26.31	13.03	33.19	11.38	40.10	11.09	43.85	8.78
	(*N* = 10)		(*N* = 14)		(*N* = 17)		(*N* = 13)		(*N* = 22)		(*N* = 17)		(*N* = 7)	
Digit Span	4.00	2.26	6.78	2.19	7.11	3.21	8.46	2.25	9.40	2.59	8.50	2.47	8.42	1.13
Mazes	13.10	5.08	14.35	5.82	15.76	5.79	21.46	3.50	19.31	4.65	21.35	4.30	23.16	4.35

items are presented in the tables in the same sequence as they appear on the WISC-R. Tabulation of the data from the performance subtests, with the exception of coding, has been omitted in the interest of parsimony.

Of the 167 subtest items tested, 157 were answered by at least one subject. From these 157 point biserial correlation coefficients, 95 are significantly different from 0 at the .05 level (approximately 60%). This moderate percentage of items showing significant validity is due in part to the limited age range sampled. In the distribution of item difficulties for 157 scorable items attempted by this population, 39 or approximately 25% of the items are within the difficulty range of .30 to .70 which is considered desirable from a test construction point of view (Anastasi, 1976). Sixty-eight of the 157 scorable items are in the extreme deciles, i.e., upper and lower ten percentiles (50 of 104 verbal items and 18 of 53 performance items). These items at the extreme deciles do not contribute significantly to the total test variance.

TABLE 2
FREQUENCY OF CORRECT RESPONSE, INDEX OF DIFFICULTY, AND INDEX OF DISCRIMINATION FOR ITEMS ON INFORMATION

Item	*f*	*p*	*rpb*	Item	*f*	*p*	*rpb*
1	172	.97	.06	16	4	.02	.04
2	174	.98	.04	17	3	.01	.03
3	171	.96	.07	18	10	.05	.09
4	161	.91	.16	19	9	.05	.10
5	124	.70	.53	20	5	.02	.05
6	119	.67	.50	21	3	.01	.03
7	110	.62	.65	22	2	.01	.02
8	110	.62	.62	23	0	.00	.00
9	98	.55	.47	24	1	.006	.01
10	98	.55	.65	25	2	.01	.02
11	63	.35	.59	26	1	.006	.01
12	26	.14	.29	27	1	.006	.01
13	33	.18	.23	28	0	.00	.00
14	19	.10	.21	29	0	.00	.00
15	18	.10	.20	30	0	.00	.00

TABLE 3
FREQUENCY OF CORRECT RESPONSE, INDEX OF DIFFICULTY, AND INDEX OF DISCRIMINATION FOR ITEMS ON SIMILARITIES

Item	*f*	*p*	*rpb*
1	151	.84	.19
2	125	.70	.50
3	104	.58	.42
4	123	.69	.46
5	127	.71	.48
6	111	.62	.52
7	95	.52	.61
8	93	.50	.66
9	74	.41	.39
10	15	.08	.16
11	12	.06	.13
12	9	.05	.10
13	5	.02	.05
14	4	.02	.04
15	6	.03	.06
16	3	.01	.03
17	1	.006	.01

Subtest Item Placement

Information. The data shown in Table 2 indicate some inconsistency of item placement relative to the WISC-R format. Items 13, 18, and 19 appear to be misplaced in order of difficulty. Item 13 (stomach) was demonstrated to be easier than item 12 (America). Items 18 and 19 are easier than items 15 and 16 for this population. The American history content of items 15 and 16 may be responsible for this finding as was demonstrated by Vernon (1977) using the WISC-R with Canadian children. A substantial increase in difficulty was noted from item 11 (seasons) through the remaining items of the test.

TABLE 4

FREQUENCY OF CORRECT RESPONSE, INDEX OF DIFFICULTY, AND INDEX OF DISCRIMINATION FOR ITEMS ON ARITHMETIC.

Item	*f*	*p*	*rpb*
1	170	.96	.09
2	171	.96	.07
3	164	.93	.14
4	147	.83	.34
5	157	.88	.19
6	156	.88	.22
7	141	.79	.32
8	111	.62	.59
9	99	.55	.61
10	52	.29	.47
11	65	.36	.50
12	48	.27	.44
13	28	.15	.31
14	26	.14	.27
15	4	.02	.04
16	6	.03	.06
17	1	.006	.00
18	1	.006	.00

TABLE 5

FREQUENCY OF CORRECT RESPONSE, INDEX OF DIFFICULTY, AND INDEX OF DISCRIMINATION FOR ITEMS ON VOCABULARY

Item	*f*	*p*	*rpb*	Item	*f*	*p*	*rpb*
1	165	.93	.14	17	61	.34	.37
2	167	.94	.12	18	32	.18	.32
3	167	.94	.12	19	6	.03	.06
4	162	.91	.15	20	7	.04	.07
5	170	.96	.06	21	1	.006	.01
6	161	.91	.12	22	1	.006	.01
7	128	.72	.29	23	3	.01	.03
8	134	.75	.40	24	1	.006	.01
9	142	.80	.29	25	1	.006	.01
10	126	.71	.40	26	2	.01	.02
11	128	.72	.29	27	0	.00	.00
12	82	.46	.51	28	0	.00	.00
13	72	.40	.38	29	0	.00	.00
14	45	.25	.28	30	1	.006	.01
15	25	.14	.28	31	0	.00	.00
16	19	.10	.21	32	0	.00	.00

Similarities. With the exception of item 3 (shirt-hat), placement according to difficulty was found to be consistent with the WISC-R. There was a sharp decline in correct responses after item 10 indicating that items following contributed little to the total subtest discriminative ability for this population.

Arithmetic. The difficulty level of items in this subtest closely approximates the order appearing in the WISC-R manual. Only items 4 and 10 appear to be misplaced with respect to ascending order of difficulty. If item 4 were placed in

TABLE 6
FREQUENCY OF CORRECT RESPONSE, INDEX OF DIFFICULTY, AND INDEX OF DISCRIMINATION FOR ITEMS ON COMPREHENSION

Item	*f*	*p*	*rpb*
1	162	.91	.11
2	152	.85	.27
3	154	.87	.24
4	152	.85	.09
5	138	.78	.33
6	124	.70	.29
7	96	.54	.42
8	71	.40	.41
9	63	.35	.55
10	39	.23	.10
11	24	.14	.25
12	21	.12	.22
13	7	.04	.07
14	8	.04	.09
15	7	.04	.07
16	6	.03	.06
17	0	.00	.00

second position it would not interfere with items 2 and 3 which require the child to cover up a number of trees.

Vocabulary. Item placement for this population is at variance with the arrangement found in the WISC R manual, particularly for items 5, 9, 17 and 18. The following was found to be more accurate in terms of ascending order of difficulty for this population: 5, 2, 3, 1, 4, 6, 9, 8, 7, 11, 10, 12, 13, 17, 14, 18, 15 and 16. An abrupt increase in difficulty from item 15 on was noted.

Comprehension. Inspection of the data in Table 6 indicates an order of difficulty closely approximating the WISC-R manual. Exchange of items 2 and 3 for this population would appear to be warranted. An abrupt shift in difficulty level is noted from item 6 to item 8.

Picture Completion. The arrangement of items on this subtest suggests that items 9 (ladder) and 10 (dresser) are misplaced in difficulty for this sample. The range of scores for this subtest indicates, however, that the discriminatory power for the majority of items is quite strong.

Picture Arrangement. For this sample, only item 4 is misplaced with respect to ascending order of difficulty. Because of the item's high discriminatory power, the placement error is not serious. There is a gradual shift from items which appear to be quite easy to items which appear relatively difficult, indicating a good range of scores. Most items contribute to the total score.

Block Design. The first item of this subtest appears to be misplaced in regard to difficulty for the present sample. It is possible that very young children have difficulty in following examples in this subtest. Also, a child's score is affected by the orientation of blocks. The last two items contribute little to the discriminatory power of the subtest for this sample.

Object Assembly. Placement of items in order of difficulty is not as important for this subtest since all designs are given. Almost all subjects achieve some level of

TABLE 7
FREQUENCIES OF SCORES TO CODE A AND CODE B

Code A		Code B			
Score	Frequency	Score	Frequency	Score	Frequency
8	1	1	1	37	6
11	1	3	1	38	4
13	1	14	1	39	5
15	2	16	1	40	6
17	1	17	3	41	3
18	3	18	1	42	2
20	1	20	3	43	4
21	1	21	2	44	2
22	2	22	3	45	2
24	1	23	5	46	2
26	1	24	4	47	2
27	4	25	4	48	2
29	1	26	3	49	1
30	2	27	3	51	2
31	3	28	3	53	1
34	1	29	3	54	1
35	2	30	2	55	1
36	3	31	5	57	1
38	1	32	5	58	3
44	1	33	4	59	1
46	1	34	5	62	1
48	1	35	2	64	1
		36	8		
	Σ = 40				Σ = 125

success on all items. For this sample, item 2 (horse) was more difficult than the others.

Coding. Since this subtest does not present an increasing item difficulty component, the results are presented in terms of frequency with which scores of Coding A (for 6- and 7-year-olds) and Coding B (for children 8 years and older) were demonstrated in this sample (Table 7).

Scaled Scores for Native Indian Children

Scaled score equivalents were calculated on the raw score distribution for each subtest and for each age in the sample. In order to retain comparability with the traditional use of the test, a mean of 10 and a standard deviation of 3 were used for each subtest. The scaled scores for the subtests can thus be added as instructed by the Wechsler manual and the sums converted into I.Q. equivalents.[1] This procedure allows the average Native Indian child of a given age to obtain an I.Q. of 100. For a quantitative interpretation of the relationship between I.Q., scaled scores, and deviation from the mean, see Wechsler, 1974, Chapter 3.

Discussion

The present study examined the performance of Native Indian children on the WISC-R. Previous findings were confirmed insofar as these children as a group displayed average perceptual motor skills and poor verbal skills. However, of greater significance for clinical interpretation were the large number of items which did not contribute significantly to the total test variance. A number of factors may be

responsible for this phenomenon. First, as previously mentioned, the age range of children sampled (6 to 13 years) represents an attenuation of the full age range for which the test was constructed. Second, the items dealing with American history which were found to be invalid for the Canadian population by Vernon (1977) are no doubt equally inappropriate for the present sample. Vernon's substitute items are therefore recommended for use with Canadian Native Indian children. Third, the Native Indian children in the present sample live on reserves where indigenous cultural values most likely prevail. Competitive aspects of life (particularly of an academic nature), orientation toward speed, accuracy, and verbal fluency play a minor role in Native Indian culture. Familiarity with the spoken and written word is compounded by a degree of bilingualism as the elders frequently converse in their native language. A further factor to be considered is the extent to which marginal socioeconomic status is evident among the Native Indian population. This variable was not formally evaluated in the present study; however, seasonal fishing is known to be the prime source of income for the Bands sampled. The values and attitudes reflected on the verbal subtests of the WISC-R, particularly Information, Comprehension, and Vocabulary are still those of the White middle class population.

The implications of the above findings are that clinicians using the WISC-R with Native Indian children should interpret their findings with extreme caution. The norms developed from the present sample are based on small numbers and a limited geographical region (Northern Vancouver Island). They should therefore be used with these factors and the previously discussed findings in mind. Because of the poor item validities reported, especially on verbal subtests, individual profile analysis as a tool of clinical inference should be avoided.

This study has not solved the clinician's dilemma in dealing with psychometric assessment of ethnic groups such as the Native Indian population. However, it has pointed out the dangers inherent in employing the WISC-R and using the norms developed on a large number of persons representative of the general population.

Notes

1. The raw scores and scaled score equivalents across all ages of the sample are available on request from the first author.

This project has been funded by the Glendale Lodge Society, Victoria, British Columbia. The testing was carried out by V. Lahmer and R. Tucker. The authors are grateful to Ms. N. Nelson for many valuable discussions and suggestions.

References

Anastasi, A. *Psychological testing.* New York: MacMillan, 1976.

Brophy, W. A., & Aberle, S. D. *The Indian: America's unfinished business.* Norman, OK: University of Oklahoma Press, 1966.

Carleton, F., & Stacey, C. An item analysis of the Wechsler Intelligence Scale for Children. *Journal of Clinical Psychology,* 1955, *11,* 149-154.

Cazden, C., & John, V. Learning in American Indian children. In M. Wax, S. Diamond & F. Gearing (Eds.), *Anthropological perspectives on education.* New York: Basic Books, 1971.

Downing, J., Ollila, L., & Oliver, P. Cultural differences in children's concepts of reading and writing. *British Journal of Educational Psychology,* 1975, *45,* 312-316.

Droege, R. C., Botterbusch, K. F., & Hawks, S. A. *Relationship between cultural exposure and attitude.* Experimental Publication Systems, 1969, 3, No. 098c.

Gaddes, W. H., McKenzie, A., & Barnsley, R. Psychometric intelligence and spatial imagery in two Northwest Indian and two White groups of children. *Journal of Social Psychology,* 1968, *75,* 35-42.

Goodenough, F. Racial differences in the intelligence of school children. *Journal of Experimental Psychology,* 1962, *9,* 388-397.

Guion, R. M. Employment tests and discriminatory hiring. *Industrial Relations,* 1966, *5,* 20-37.

MacArthur, R. S. Assessing intellectual potential of Native Canadian pupils. *Alberta Journal of Educational Research,* 1968, *14* (2), 115-122.

MacArthur, R. S. Some cognitive abilities of Eskimo, White and Indian Metis pupils aged 9 to 12 years. *Canadian Journal of Behavioural Science,* 1969, *1* (1), 50-59.

Peters, H. Performance of Hopi children on four intelligence tests. *Journal of American Indian Education,* 1963, *2,* 27-31.

Rohrer, J. The test intelligence of Osage Indians. *Journal of Social Psychology,* 1942, *16,* 99-105.

Spreen, O., & Tryk, H. E. WISC informaton subtest in a Canadian population. *Canadian Journal of Behavioural Science,* 1970, *2,* 294-298.

Taylor, L. V., & Skanes, C. Cognitive abilities in Inuit and White children from similar environments. *Canadian Journal of Behavioural Science,* 1976, *8* (1), 1-8.

Vance, H., & Gaynor, P. A. Note on cultural difference as reflected in the WISC. *Journal of Genetic Psychology,* 1976, *129,* 171-172.

Vernon, P. E. Educational and intellectual development among Canadian Indians and Eskimos. Part 1, 2. *Educational Review,* 1966, *18,* 79-91, 186-196.

Vernon, P. E. WISC-R. *Canadian Psychological Association Bulletin,* 1974, *4,* 8-9.

Vernon, P. E. Modifications of the WISC-R for Canadian use. *Canadian Psychological Association Bulletin,* 1976, *6* (1), 4-5.

Vernon, P. E. Final report on modifications of WISC-R for Canadian use. *Canadian Psychological Association Bulletin,* 1977, *5* (1), 5-7.

Wechsler, D. *Manual for the Wechsler Intelligence Scale for Children — Revised.* New York: Psychological Corporation, 1974.

Wilson, L. Canadian Indian children who never attended school. *Alberta Journal of Educational Research,* 1973, *19* (4), 309-313.

The Alberta Journal of Educational Research *Vol. XXVI, No. 1, March, 1980*

R. G. HACKER
The University of Western Australia

Cognitive Interactions in Science: Classroom Practices and some Prescriptions of Theories of Learning

An experimental analysis of the cognitive interactions of 320 lessons taught by 54 science teachers at 6 secondary grade levels is reported. Variance in classroom practices was explained in terms of a number of dimensions of cognitive teaching style.

From a study of changes in these dimensions across the grade levels, it was concluded that observed practices were generally harmonious with the notion of a hierarchical progression from knowledge of facts and principles to applications and problem-solving activities. However, low levels of practical activity in the early secondary grades revealed considerable dissonance between classroom practices and the prescriptions of Piagetian theory.

During the past decade classroom interaction analysis has become firmly established as a distinctive area of educational research. From rather chaotic beginnings, a systematic overview is now emerging of the types of studies which have been undertaken in science classrooms.

Power (1977) provides an interesting map of the territory of science classroom interaction studies and he suggests a useful classification of studies, according to the selected aspects of classroom life upon which attention is focussed. Affective studies are those attempting to characterize the socio-emotional climate of the classroom, whereas managerial studies focus on organizational, procedural and disciplinary aspects of the classroom processes. Cognitive studies, which seek to describe the intellectual transactions taking place in the classroom, form a third category of Power's classification.

In retrospect, it is perhaps unfortunate that the great majority of the early

---

Dr. Hacker has undertaken graduate studies in both science and education at the Universities of Waterloo and London, respectively. He has taught at the secondary and tertiary levels on three continents and is currently with the Department of Education, The University of Western Australia. Dr. Hacker's current research interests include classroom interaction analysis studies and strategies for curriculum diffusion.

studies undertaken tended to focus rather narrowly on either the socio-emotional climate or on specific managerial issues. Many of these studies employed Flanderian-type observation systems (Flanders, 1970) or extensions and modifications of these systems. Of course it cannot be denied that effective class management and a healthy emotional climate are essential prerequisites to learning, but they are not sufficient. Given adequate teacher management skills and a reasonable rapport between the teacher and the class, it can be argued strongly that the effectiveness of science learning and teaching must depend ultimately on the quality of the intellectual transactions which occur, that is, upon the types of cognitive skills which students are able to practise in the classroom.

Surprisingly few empirical studies of the cognitive aspects of science teaching have been undertaken and there is a clear need for more descriptive and correlational studies to tell us how classrooms do vary and how these variations might relate to the cognitive development of students in science.

This study aims, via an experimental analysis of classroom behaviours, to identify and quantify dimensions along which cognitive interactions vary from one classroom to the next. Also, by examining trends in variations across different grade levels, the study aims to provide a basis for the comparison of classroom practices with some prescriptions of theories of learning.

A review of relevant educational and psychological theory provides a number of potentially useful distinctions between different kinds of cognitive behaviours, upon which an observational instrument might be built. For example, one dimension along which the cognitive behaviours found in science classrooms might vary is Guildford's (1956) convergent/divergent dimension. Another distinction, provided by the work of Bloom (1956), Gagné (1970) and others, is the notion of a hierarchical progression from knowledge of simple facts and principles to applications and problem solving activities. Flanders (1970) delineates teacher-dominated interactions and those interactions which are initiated by students. Armstrong (1910), Bruner (1960), Schwab (1962) and others differentiate between didactic and heuristic interactions and, in addition, the studies of Inhelder and Piaget (1958) emphasize the importance of a progression from concrete, experimental experience to more abstract, theoretical understanding.

This review identifies a number of different dimensions along which the cognitive interactions of science lessons are likely to vary. It also suggests that an adequate description of classroom practices is more likely to result from the development of multidimensional profiles than from the use of vague and global dichotomies, such as "traditional" and "progressive" which have previously been used to describe teaching.

It is clear that the selection or development of an observational instrument for this type of study merits careful consideration and Rosenshine and Furst (1973) provide a useful series of suggestions to aid the would-be classroom researcher in this decision-making process. Many of the earlier systems which were developed were rejected because they lacked any explicit or implicit theoretical or empirical base. Indeed many of these systems seem to have been developed from little more than the author's fantasies about classroom life. Observational instruments based on one established educational or psychological theory were also considered to be unsuitable because of the restrictions which this framework would inevitably impose on the data which were subsequently recorded. Instead, as outlined above, a multidimensional instrument was sought, which would facilitate recording of a very

broad range of those cognitive behaviours which might be expected to occur in science lessons.

Before developing a new instrument and adding, perhaps needlessly, to the great array of systems which have been developed for observing classrooms, a search was undertaken to see if any existing system might be located which would serve the purposes of this inquiry. The main selection criteria would be, firstly, that the instrument should provide quantitative measure of the frequencies of a wide range of cognitive behaviours and, secondly, that the instrument should have been specifically developed to study science lessons. The latter criterion was introduced because general systems have often failed to cope adequately with the unique aspects of science lessons in a laboratory context.

Eventually the Science Teaching Observation Schedule (STOS), developed by Eggleston, Galton and Jones (1975), was selected as being well-suited to the requirements of this study. The STOS instrument was developed initially from an empirical analysis of videotaped science lessons and was specifically designed to differentiate between different styles of cognitive interaction in the science classroom. It is concerned only with the cognitive dimensions of classroom interactions and does not seek to quantify either socio-emotional or managerial aspects of classroom communication.

The design of the schedule is based on two explicit assumptions: firstly, that recent developments in science teaching accept the processes of science as an integral part of science learned in schools and, secondly, that student cognitive growth in science is determined largely by the opportunities which are provided in the classroom to practice the various intellectual skills which are associated with science and scientific inquiry.

The 23 categories of the STOS instrument are shown in Table 1. To facilitate the recording of classroom events, observations are grouped according to those events which are initiated by the teacher and those events which are initiated or maintained by students. Teacher talk is divided into three major areas: teacher questions, statements, and directives; pupil activity is divided into two main areas: pupil seeks information or consults and pupils refer to the teacher. Each of these five main areas is subdivided into minor categories, the basic pattern being interactions which are associated with recall of facts and principles, applications to problem-solving, hypothesising and experimental procedures. One interesting feature of the teacher questioning categories is that the observer's attention is focussed on the responses actually made by the pupils and those cognitive skills which they are practising. In this context, problem-solving is defined in a convergent sense where only one solution to the problem is acceptable to the teacher.

This examination of the STOS categories shows that the instrument has the potential to provide simultaneous measures along a number of different dimensions of science classroom interactions. A speculative model, which relates to the assignment of STOS categories to three major dimensions of science teaching, has been developed by Eggleston et al. (1976). This model is shown in Figure 1 and gives the authors' speculative assignment of the categories to three scales, represented as 3 orthogonal dimensions.

It is apparent that categories $a_{1\text{-}7}$, $b_{1\text{-}4}$ and $c_{1\text{-}4}$ must, by definition, be teacher-directed or dominated and, conversely, that categories $d_{1\text{-}4}$ and $e_{1\text{-}4}$ must be pupil-directed. However, the arguments given for assignment of categories along

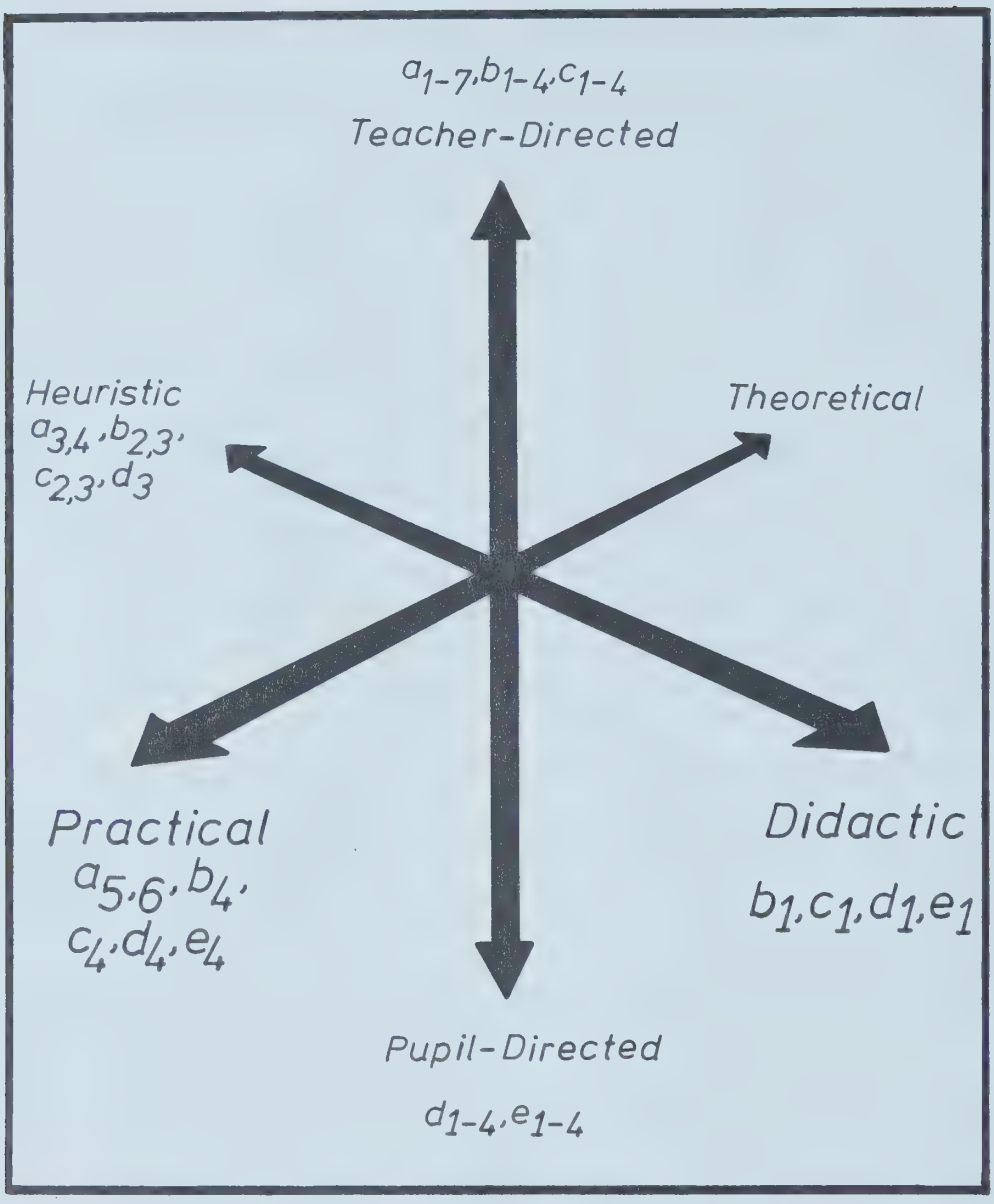

Figure 1. Speculative assignment of the STOS categories to a 3-dimensional model of science teaching.

the practical-theoretical and heuristic/didactic dimensions are more speculative. By exploring underlying dimensionalities in observed classroom interactions, this study provides an opportunity to test whether the internal structure of classroom observations is adequately summarized by this model and, more specifically, to check the assignment of the STOS categories to the various dimensions.

Method

Sample

The sample consisted of 54 teacher-class units from high schools across the three Maritime provinces of Canada. Nine teacher-class units for each of the grade levels 7 to 12 participated in the study and equal numbers of classes studying biology, chemistry, and physics topics, respectively, were included in this sample. Classes were generally unstreamed. At schools where classes were streamed at the higher grade levels, a class from the middle ability in science was selected.

Instrument

The categories incorporated into the STOS instrument were described earlier. A sign system for recording data was used, with a time-sampling unit of three minutes. Each category was checked once if it occurred during this period. A user's manual for the instrument and well developed training procedures for observers are available (Eggleston et al., 1975).

Procedure

Before the observational work was started, some pilot studies were carried out to see if any modifications to the STOS instrument would be necessary for use in Canadian schools. Observer reactions to the instrument were favourable and it was thought that valid distinctions between different cognitive teaching styles could be made with the standard instrument.

Careful attention was given to the issue of observer reliability. The classroom observation work was preceded by a full week of intensive training in use of the STOS instrument. At the end of this week reliability trials were carried out according to procedures recommended by Eggleston et al., (1975). During observation periods in the schools further reliability tests were given and retraining sessions were held as required to maintain high levels of observer reliability.

Contacts with schools were made via the superintendent of the school district. Principals were asked to forward the names of science staff who were willing to participate in the study. These teachers were asked to select six lessons which they considered to be typical of their teaching; visits to classes by different observers were randomized to avoid systematic observer errors in the results.

Results

For each teacher-class unit a percentage use figure was calculated for each of the 23 STOS categories. For example, if a certain teacher was observed for a total of 60 three-minute, time-sampling units and use of category a_1 was recorded in 45 of those units, then that teacher's use of category a_1 was recorded as 75 percent.

To identify those underlying dimensions along which cognitive interactions do vary, a principal components analysis of this frequency data with a varimax rotation was carried out. Before undertaking this analysis, 4 infrequently-used STOS behaviours (a_4, a_7, c_3, and d_3) were discarded and the remaining 19 variables were subjected to a log transformation. This treatment was given to ensure that the distributions of input data for the factor analysis met the requirements of approximating to nonsingular multinormality.

Five factors with eigen values in excess of unity were extracted from this analysis. These factors are described below. A varimax rotation of the 19 STOS behaviours is shown in Table 1.

Factor 1 provides a measure along the knowledge/applications dimension of science teaching. It is a "problem-solving" factor which relates to a shift in emphasis in the classroom from knowledge of facts and principles to applications and problem-solving activities.

Factor II relates to the familiar practical/theoretical dimension of science teaching. It is a "practical" factor which provides a measure of the amount of class time spent on demonstrations and class experiments.

Factor III is concerned with the didactic/heuristic dimension of science teaching. It

TABLE 1

STOS CATEGORIES AND VARIMAX ROTATED FACTOR LOADINGS FOR THE SAMPLE OF SCIENCE TEACHERS ($N = 54$)

		I	II	III	IV	V
1.	TEACHER TALK					
1-a	Teacher asks questions (or invites comments) which are answered by:					
	a_1 recalling facts and principles	-33				
	a_2 applying facts and principles to problem solving	+70				
	a_3 making hypothesis or speculation			+79		
*	a_4 designing of experimental procedure					
	a_5 direct observation		+61			+85
	a_6 interpretation of observed or recorded data			+67		
*	a_7 making inferences from observations or data					
1-b	Teacher makes statements:					
	b_1 of fact and principle	-27			+73	
	b_2 of problems	+86				
	b_3 of hypothesis or speculation			+86		
	b_4 of experimental procedure		+90			
1-c	Teacher directs pupils to sources of information for the purpose of:					
	c_1 acquiring or confirming facts of principles	-31			+79	
	c_2 identifying or solving problems	+87				
*	c_3 making inferences, formulating or testing hypotheses					
	c_4 seeking guidance on experimental procedure		+85			
2.	TALK AND ACTIVITY INITIATED AND/OR MAINTAINED BY PUPILS					
2-d	Pupils seek information or consult for the purpose of:					
	d_1 acquiring or confirming facts or principles	-21			-90	
	d_2 identifying or solving problems	+88				
*	d_3 making inferences, formulating or testing hypotheses					
	d_4 seeking guidance on experimental procedure		+87			
2-e	Pupils refer to teacher for the purpose of:					
	e_1 acquiring or confirming facts or principles	-42				
	e_2 seeking guidance when identifying or solving problems	+84				
	e_3 seeking guidance when inferring, formulating or testing hypotheses			+46		
	e_4 seeking guidance on experimental procedure		+90			
	% of variance	26.8	18.0	11.4	7.8	5.5

* Infrequently used categories which were excluded from the factor analysis.

is a "heuristic" factor which relates to the degree to which science is presented as a speculative and exploratory activity.

Factor IV appears to relate to the teacher-dominated/pupil-initiated dimension, concerning the degree to which the initiative is held by the teacher in the classroom.

Factor V is a teacher-questioning factor relating to the development of observational skills in pupils.

For each teacher, a score for each of these 5 factors was calculated. These factor-scores were calculated from a direct summation of the logs of the percentage use figures rather than from a weighting of factor-score coefficients.

A two-way multivariate analysis of variance of this data was carried out with subject areas (biology, chemistry, and physics) and grade levels (7 to 12) as

TABLE 2

SUMMARY OF THE MULTIVARIATE ANALYSIS OF VARIANCE DATA

Source	Multivariate F	Factor	Univariate F
Subject Areas			
Biology v (Chemistry + Physics)	9.77***	I	48.81***
		II	.49
		III	1.97
		IV	0.15
		V	3.24
Chemistry v Physics	1.89	I	1.76
		II	2.03
		III	2.11
		IV	1.11
		V	3.53
Grade Levels			
Linear	20.61***	I	34.98***
		II	10.02**
		III	1.78
		IV	39.60***
		V	3.48
Quadratic	3.00*	I	1.40
		II	5.03*
		III	0.52
		IV	6.54*
		V	0.08
Cubic	6.94***	I	5.01*
		II	6.34*
		III	0.17
		IV	22.41***
		V	1.72
OTHERS	-	-	-
INTERACTION	-	-	-

* $p < .05$
** $p < .01$
*** $p < .001$

independent variables. Previous research work (Eggleston et al., 1976) had pointed towards significant differences between biological and physical science (chemistry and physics) lessons. Consequently, Helmert contrasts were applied to levels of the subject area variable. Orthogonal polynomial contrasts were applied to levels of the grade level variable. The results of this analysis are provided in Table 2. Figure 2 shows mean Factor I scores for biological and physical science lessons at different grade levels. Figure 3 shows trends in Factors I, II and IV across the grade levels for all science lessons. In addition the percentage of teacher-dominated interactions (categories $a_{1\text{-}7}$, $b_{1\text{-}4}$, and $c_{1\text{-}4}$) was calculated for each teacher. Mean values across the grade levels are shown in Figure 4.

Discussion

The source variables which emerged from the principal components analysis were readily interpreted in terms of recognizable and familiar dimensions of cognitive style, providing encouraging support for the speculative model proposed by Eggleston et al. (1976).

Factors II, III and IV were directly related to the three main dimensions of the model, accounting for almost 40 percent of the variance between teachers. In addition, Factor I, relating to a shift in emphasis from knowledge of facts and principles

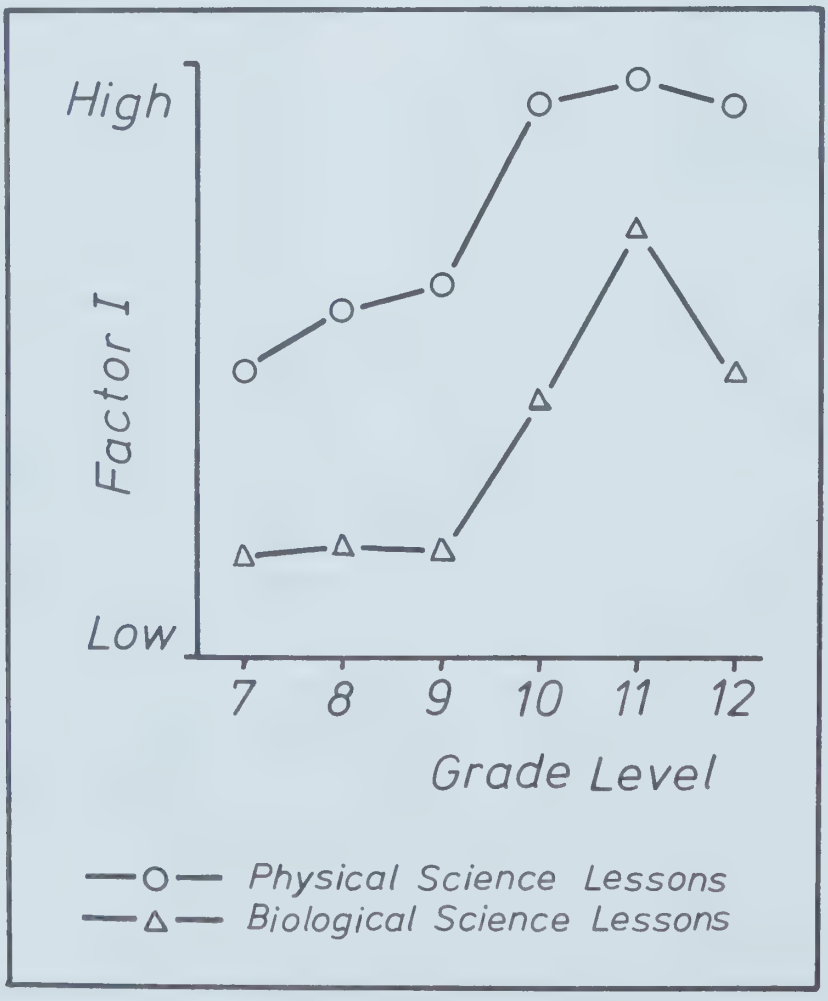

Figure 2. Mean Factor I scores for biological and physical science lessons across the grade levels.

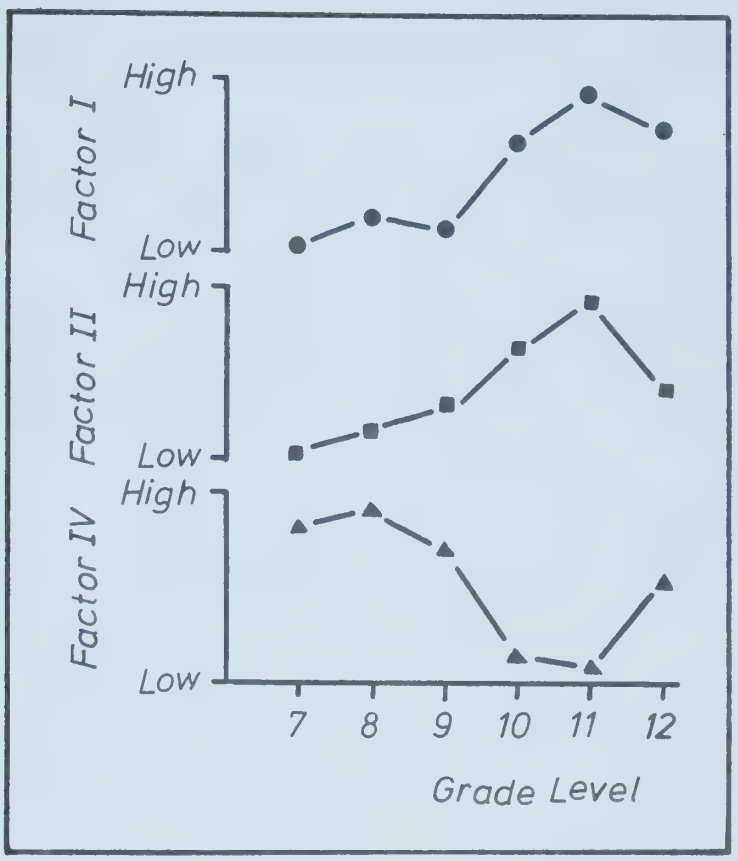

Figure 3. Trends in Factors I, II, and IV across the grade levels for all science lessons.

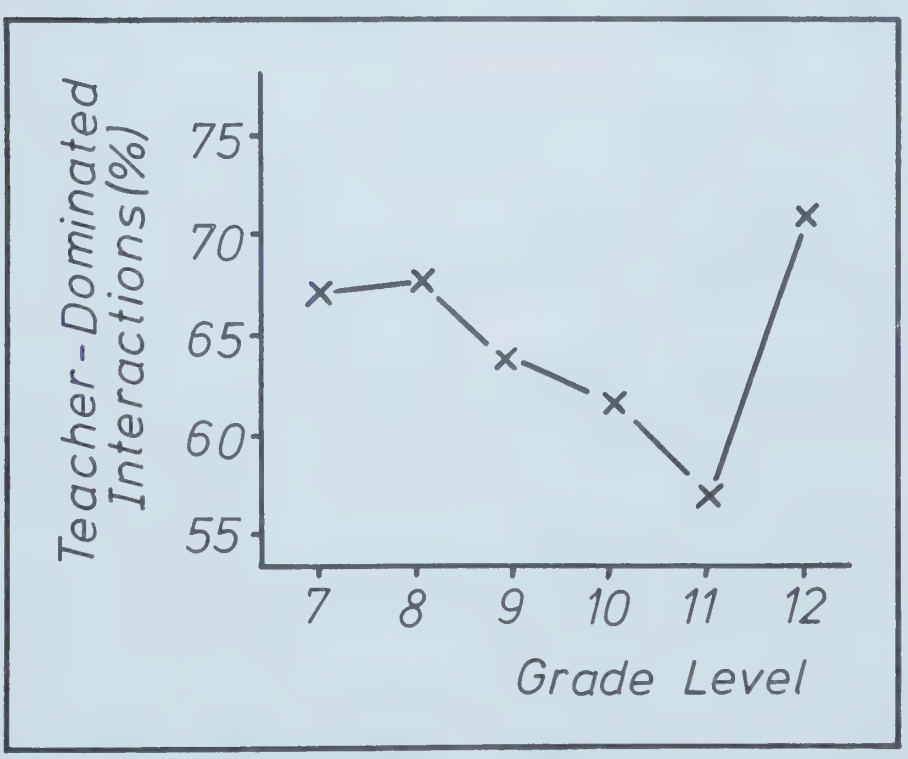

Figure 4. Mean percentages of teacher-dominated interactions across the grade levels for all science lessons.

to applications and problem-solving activities, accounted for more than a quarter of the variance in cognitive style. This fourth dimension is proposed as a valuable extension to the basic model.

With regard to the three dimensions of the basic model, some minor relocations of categories along the axes might be suggested. In particular, the association of category a_6 with the practical end of the practical/theoretical axis might be questioned. Questions answered by interpretation of observed or recorded data would seem to be more closely linked with heuristic teaching strategies. Indeed, observers commented that, during most practical lessons, teacher questions answered by observation were more usually followed by a teacher statement of fact or principle (category b_1) rather than an invitation to interpret these observations. Also the association of categories b_2 and c_2 with heuristic teaching strategies may be weaker than the speculative model suggests.

A further interesting point is that questions answered by direct observation would seem to occur in two different contexts. The first is the more common association with experimental work in a laboratory, while the second context, which probably accounts for the emergence of Factor V, is in association with materials, samples, or photographs brought to the lesson by students or teachers.

Some evidence for the validity of the interpretation of Factor IV as a measure of teacher dominance was provided by the plot of the more familiar Flanderian index given in Figure 4, insofar as variations in this index and Factor IV across the grade levels were quite similar.

The multivariate analysis of variance data (see Table 2) confirmed significant variation in one dimension to the science discipline being taught. Biology lessons were differentiated from physical science lessons by differences in Factor 1 scores. As shown in Figure 2, at all grade levels, lower scores for biology lessons point towards a greater emphasis on facts and principles, whereas in physical science lessons, higher scores indicate more emphasis on applications and problem-solving activities.

Across the grade levels, significant trends in Factors I, II and IV were found (see Table 2 and Figure 3). It is interesting to note that no significant variations in Factor III were found, either with subject taught or grade level. The degree to which science is presented as a speculative activity would seem to be more closely linked with the individual teacher's approach rather than with these discipline or pupil variables.

For all science subject areas, Factor I scores increased across the grade levels, with a reversal of this trend, particularly in biology classes, at the grade 12 level. The general trend through the grade levels is harmonious with Gagné's (1970) notion of cognitive development in science following a hierarchical progression from knowledge of simple facts and principles through to applications of these facts and principles to problem-solving activities.

The trend in Factor II (see Figure 3) revealed considerable dissonance between classroom practices and the prescriptions of Piagetian theory. At the early secondary grade levels very little practical work is undertaken, yet Piaget's observations and classifications of child development clearly indicate the need for concrete, practical experience at this time to encourage and facilitate learning. The general trend is towards increased amounts of experimental work at the higher grade levels. This trend is also reversed during the final year of schooling.

From a theoretical perspective, the picture of science education at the early secondary grade levels which emerges from this study is a disturbing one. High levels of teacher-dominance and low levels of practical activity in the early years are thought not to provide a sound basis for cognitive development in science. Indeed, there is substantial evidence (Duckworth, 1964; Green, 1971) to suggest that deprivation of practical experience at this stage can actually retard further cognitive growth.

In conclusion, the sharp reversal of all observed trends during the final year of secondary schooling was an unexpected finding. These changes reflect a return to teacher-directed teaching strategies with renewed emphasis on knowledge of basic facts and principles rather than applications to problem-solving activities. Lungdren (1972) has proposed a model of teaching whereby it is viewed as a process governed by "frames" or externally imposed constraints which determine what is possible or not possible in the classroom. It seems likely that the changes observed during the final year may be linked with external factors, such as teachers' perceived needs to complete a course of study, but further research activity is obviously needed to establish reasons for these variations in patterns of classroom interactions.

References

Armstrong, H. E. *The teaching of scientific method* (2nd ed.). London: MacMillan, 1910.

Bloom, B. S. (Ed.). *Taxonomy of educational objectives, Handbook I.* New York: McKay, 1956.

Bruner, J. S. *The process of education.* Harvard University Press, 1960.

Duckworth, E. Piaget rediscovered. *Journal of Research in Science Teaching,* 1964, *2,* 172-175.

Eggleston, J. F., Galton, M. J., & Jones, M. E. *A science teaching observation schedule.* London: Macmillan Education, 1975.

Eggleston, J. F., Galton, M. J., & Jones, M. E. *Processes and products of science teaching.* London: Macmillan Education, 1976.

Flanders, N. A. *Analyzing teacher behaviour.* New York: Addison-Wesley, 1970.

Gagné, R. M. *The conditions of learning (2nd ed.).* New York: Holt, Rinehart, and Winston, 1970.

Green, D. R. (Ed.). *Measurement and Piaget.* New York: McGraw-Hill, 1971.

Guildford, J. P. The structure of intellect. *Psychological Bulletin,* 1956, *53,* 267-293.

Inhelder, B., & Piaget, J. *The growth of logical thinking from childhood to adolescence.* London: Routledge and Kegan Paul, 1958.

Lungdren, U. *Framing factors and the teaching process.* Goteborg: Almqvist and Wiksell, 1972.

Power, C. A critical review of science classroom interaction studies. *Studies in Science Education,* 1977, *4,* 1-25.

Rosenshine, B., & Furst, N. The use of direct observation to study teaching. In R. W. Travers (Ed.), *Second handbook of research on teaching.* Chicago: Rand McNally, 1973.

Schwab, J. J. *The teaching of science.* Harvard University Press, 1962.

EDWARD G. SUMMERS
The University of British Columbia

The Validity of the *Estes Reading Attitude Scale* for Intermediate Grades

Affective development is being increasingly emphasized in learning, but problems arise in obtaining good measures of attitude toward school subjects, including reading. This study examines the internal consistency reliability and convergent validity of the widely used Estes Reading Attitude Scale *(ERAS) for intermediate grades. Subjects were 1,403 children in 65 intact classrooms in nine elementary schools in Richmond, British Columbia. The* Metropolitan Achievement Test — Reading *(MAT), the ERAS, and two subjective measures, the* Student Reading Rating Scale *(SRRS) and the* Teacher Reading Rating Scale *(TRRS), were administered to all subjects. Results indicate good internal consistency reliability for the ERAS and solid convergent validity as indicated by the magnitude and direction of the correlations with the SRRS, the TRRS, and the MAT. The ERAS provides a useful global measure of school-oriented attitude toward reading in intermediate grades.*

Affective development is being increasingly emphasized as an educational objective. In their extensive review of a decade of research on the teaching of affective responses, Khan and Weiss (1973) concluded that "While the first wave of new curricula was concerned almost exclusively with cognitive outcomes, the most recent programs have been placing more emphasis on affective behaviours" (p. 788).

The *Taxonomy of Educational Objectives: Affective Domain* (Krathwohl, Bloom & Masia, 1964) defines the affective domain as subsuming such traits as temperament, attitude, belief, interest, or value. Attitudes, beliefs, values, and interests are important outcomes of education. Shaw and Wright (1967) point out that "If the attitude of a person toward a given object, or class of objects, is known, it can be used in conjunction with situational and other dispositional variables to

---

Dr. Edward G. Summers completed the B.S., M.A., and Ph.D. degrees at the University of Minnesota in educational psychology, reading and special education. He is currently Professor of Education, Faculty of Education, University of British Columbia where he teaches undergraduate and graduate courses in secondary reading, research, and information retrieval. His special areas of interest include measurement in reading, text structure and discourse analysis, reading and affect, and secondary, college, and adult reading instruction and research.

predict and explain reactions of the person to the class of objects" (p. 2). Indeed, with respect to reading development, Athey (1976) suggests that "by increasing our knowledge of the affective life of the child, we broaden our understanding of his total functioning in the academic situation. In other words, the intellectual variables involved in reading do not operate in isolation but are modified by the individual's attitudinal and personality characteristics" (p. 352).

Increased concern over the academic achievement of public school children necessitates a careful examination of potentially relevant factors which contribute to success in school. Since reading is the basis for achievement in most school subjects, it seems logical that if a child develops a positive attitude toward reading, the attitude may well carry over to other subjects. Questions are more frequently being asked about the affective parameters of reading, viz: Do certain practices, programs and materials in reading hinder or facilitate a positive attitude toward reading? Do children who can read, but do not, need remediation? How much do children read and why do some children read while others do not? Do all good readers have positive attitudes toward reading and/or vice versa? Answers to these questions hinge on estimating attitudes toward reading, an area of measurement which has been characterized by meager instrumentation to date in spite of its purported importance. Although attempts have been made to taxonomize affective objectives (Krathwohl et al., 1964), measurement in this area, in general, is still in its infancy in comparison to cognitive assessment and much remains to be done in defining the nature of affect, measuring it, and delineating its role in the education process (Khan & Weiss, 1973; Gephart, Engle & Marshall, 1978). A number of instruments have been developed in reading, but further work is needed to provide a solid base for assessing reading attitude (Alexander & Fillmore, 1976; Ewing, 1977; Summers, 1977; Reed, 1978).

The *Estes Scales to Measure Attitude Toward School Subjects* (Estes, 1971; Estes, Johnstone & Richards, 1975) are currently widely used in school research. The scales were developed to assess attitude in five content areas: English, mathematics, science, reading, and social studies. Validation studies thus far have shown promising results for the reading scales using subjects in grades seven through twelve (Estes & Johnstone, 1974; Estes et al., 1975) and grade eleven (Dulin & Chester, 1974). The purpose of this study is to present additional data on the internal consistency reliability and convergent validity of the *Estes Reading Attitude Scale* (ERAS) for a large population of intermediate grade subjects (Grades 5, 6, and 7).

Internal Consistency

The ERAS is a unidimensional 15-item instrument. The method of summed ratings is used in which responses are solicited to the battery of items using a five-point Likert scale with forward and reverse polarity. Subjects choose the alternative that best describes their reaction to each item. Attitude toward reading is defined as a liking for or a dislike of reading. The 15 items of ERAS are:

1. Reading is for learning but not for enjoyment.
2. Spending allowance on books is a waste of good money.
3. Reading is a good way to spend spare time.
4. Books are a bore.
5. Watching T.V. is better than reading.

6. Reading is rewarding to me.
7. Books aren't usually good enough to finish.
8. Reading becomes boring after about an hour.
9. Most books are too long and dull.
10. There are many books which I hope to read.
11. Books should only be read when they are assigned.
12. Reading is something I can do without.
13. Some part of summer vacation should be set aside for reading.
14. Books make good presents.
15. Reading is dull.

Subjects were 1,403 intermediate children in 65 intact classrooms in nine elementary schools in Richmond, British Columbia (Grade 5, $n = 448$; Grade 6, $n = 453$; Grade 7, $n = 502$). All data were collected in early December of 1977. The *Metropolitan Achievement Test — Reading* (Durost, Bixler, Wrightstone, Prescott & Balow, 1970-71) was used to determine the reading characteristics of the intermediate subjects. The mean raw scores for total reading on the Intermediate and Advanced forms placed the three grade groups at the 68th, 58th, and 50th percentiles in reading achievement. The somewhat above-average percentile ranks for the three grades are typical of obtained scores when beginning year norms are used to interpret mid-year achievement test data. On a second administration of the MAT in June, 1978, the same groups achieved at the 58th, 50th, and 50th percentiles respectively. The percentile score results and the plotted histograms for the three groups suggest a normal distribution for the subjects on the norms for the test with a slight negative skewing of scores.

Hoyt reliability estimates (Nelson, 1974) were used to determine the internal

TABLE 1
ERAS MEANS, STANDARD DEVIATIONS, STANDARD ERRORS,AND HOYT INTERNAL CONSISTENCY MEASURES BY GRADE AND GENDER

	n	mean	standard deviation	standard error	consistency coefficient
Grade 5	448	56.07	10.37	3.91	.86
Grade 6	453	55.87	10.19	3.62	.87
Grade 7	502	55.96	10.38	3.35	.89
Grade 5 male	227	53.53	10.65	3.91	.86
female	221	58.68	9.41	3.39	.86
Grade 6 male	223	54.21	10.09	3.84	.84
female	230	57.47	10.05	3.38	.88
Grade 7 male	241	53.71	10.61	3.49	.88
female	261	58.03	9.73	3.39	.88

consistency of the ERAS. (It should be noted that the Hoyt estimates are equivalent to Cronbach's [1951] alpha.) Internal consistency, an aspect of reliability sometimes referred to as homogeneity, is roughly equivalent to the average Pearson correlation between all pairs of items. Table 1 presents the means, standard deviations, standard errors, and internal consistency coefficients by grade and gender for the ERAS. The internal consistency coefficients are all beyond .80 which is very high indeed for an affective instrument (Anastasi, 1976) and compares favourably with reliabilities obtained for many cognitive measures. Internal consistency coefficients are also beyond .80 for all three grades thus supporting the construct homogeneity of the ERAS. Reliability greatly influences the size of the Standard Error of Measurement; tests with high reliability have correspondingly small SEm's. The SEm's are relatively small and quite consistent for the three grade levels thus providing further evidence of good reliability for the ERAS.

The item intercorrelation matrices (Table 2) for the three grades reveal typical patterns for a unidimensional instrument. All correlations of single items with total scores exceed .40, evidence that each specific item does a good job of measuring the construct of reading attitude (Ebel, 1972). In addition, as expected in a good unidimensional instrument, there are relatively low correlations between the separate items in the test.

Convergent Validity

Convergent (or criterion) validity can be defined as the extent to which test performance is related to some other valued measure of performance. Convergent validity is most often estimated by correlating test scores with alternative measures of the same construct. In this study, two informal instruments were developed to provide criterion measures: the *Student Reading Rating Scale* (SRRS) and the *Teacher Reading Rating Scale* (TRRS). Both measures were developed using an adaptation of the format reported by Dulin and Chester (1974). On the SRRS children are asked to rate themselves, in comparison with their peers, on three questions using a five-point Likert type scale: (1) About how *Well* do you read? (2) How well do you *Like* to read? (3) About how *Much* do you read? The range of summed scores for the scale is 3 to 15 points with a midpoint of 9. The TRRS is similar to the SRRS and asks teachers to rate children, in comparison with other children, in terms of: (1) How *Well* they read. (2) How much they *Like* to read. (3) How *Much* they read. Scoring is similar to the SRRS. Table 3 presents the means and standard deviations for the SRRS and the TRRS by grade level and Well, Like, and Much scores. The ERAS correlated .33 with the SRRS Well score combined for the three grades, .56 with the SRRS Like score combined for the three grades, and .48 with the SRRS Much score combined for the three grades. Supportive validity coefficients should be statistically significant at beyond the five percent level of confidence and of a reasonably high magnitude. There is debate on what is a "reasonably high magnitude" and recommendations in the literature range from .30 to .40 (Guilford, 1956). For our purposes, correlations of .35 and above were deemed acceptable. All the validity coefficients for the SRRS are positive, significant beyond .05 ($p < .001$), and relatively high in magnitude. Discriminant validity is also indicated as the items correlate in the expected direction; attitude correlates more highly with liking for reading and how much is read than it does with the achievement item asking how well one reads. The ERAS and SRRS validity correlations are similar in direction and magnitude to those found

TABLE 2
ERAS INTERCORRELATIONS FOR GRADES 5, 6, and 7

	Grade	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
	5	30*														
2	6	30														
	7	32														
	5	24	27													
3	6	16	32													
	7	20	38													
	5	28	29	43												
4	6	20	34	46												
	7	26	37	50												
	5	23	32	32	35											
5	6	13	27	26	31											
	7	14	40	38	41											
	5	22	31	37	42	41										
6	6	20	30	38	36	36										
	7	20	39	38	45	41										
	5	20	28	30	31	28	36									
7	6	28	22	19	26	28	24									
	7	18	31	30	42	31	28									
	5	30	22	30	37	41	37	30								
8	6	29	34	26	38	35	32	31								
	7	28	31	34	45	37	43	45								
	5	30	37	28	39	39	39	44	47							
9	6	30	37	27	43	26	41	43	48							
	7	26	31	38	51	40	39	51	54							
	5	12	22	40	35	19	35	23	22	32						
10	6	16	23	25	28	22	26	19	28	35						
	7	15	26	29	38	32	38	39	44	43						
	5	33	23	25	35	21	20	31	27	34	21					
11	6	27	29	24	26	21	23	16	23	28	27					
	7	24	29	37	40	24	30	36	36	35	34					
	5	18	24	20	34	23	23	21	19	36	27	19				
12	6	21	31	22	26	24	33	26	28	31	24	26				
	7	22	31	40	38	37	35	35	37	38	29	38				
	5	22	28	35	32	33	34	26	26	30	28	20	23			
13	6	25	31	31	33	34	33	19	32	28	26	30	21			
	7	14	27	37	29	30	35	26	26	31	27	24	24			
	5	20	23	25	25	28	32	19	25	25	23	15	19	36		
14	6	22	33	30	28	22	32	20	33	31	35	30	19	29		
	7	14	33	32	29	37	33	30	34	34	37	27	27	30		
	5	24	32	43	54	40	39	33	37	43	38	29	38	43	45	
15	6	31	41	44	45	29	44	32	44	47	40	29	41	43	42	
	7	29	37	44	61	38	39	44	49	51	41	46	43	34	40	
	5	50	56	59	67	61	64	56	61	69	52	51	50	58	50	71
Total	6	49	60	57	64	54	63	51	64	68	52	51	53	59	56	73
	7	44	61	64	71	63	65	63	69	71	61	58	60	53	57	74

*decimals omitted

by Dulin and Chester (1974) for grade eleven (r = .36, .60, .54, respectively). Estes et al. (1975) also report a correlation of .28 between student self-estimate of attitude and the ERAS.

Correlations between the ERAS and the TRRS Well, Like, and Much scores, combined for the three grades, were .36, .39, and .37, respectively. Again, although

TABLE 3
SRRS AND TRRS MEANS AND STANDARD DEVIATIONS BY GRADE

	SRRS						*TRRS*[a]					
	Grade 5		Grade 6		Grade 7		Grade 5		Grade 6		Grade 7	
	mean	sd	mean	sd	mean	sd	mean	sd	mean	sd	mean	sd
Well	3.33	.98	3.31	1.00	3.22	.95	3.11	1.05	3.07	1.14	3.13	1.13
Like	3.61	1.10	3.42	1.14	3.39	1.10	3.10	.90	3.11	1.10	3.15	1.04
Much	3.66	1.18	3.43	1.26	3.30	1.25	3.05	1.03	2.97	1.27	3.11	1.15
Total Scale	10.62	2.69	10.17	2.98	9.92	2.90	9.27	2.73	9.16	3.25	9.41	3.09
N	(448)		(453)		(502)		(448)		(453)		(502)	

[a] Based on responses from 65 teachers

of more modest magnitude, the validity correlations are positive, significant beyond .05 ($p < .001$), and in the expected direction. The ERAS and the TRRS correlations also corroborate the Dulin and Chester validity results in direction and magnitude ($r = .33, .38, .40$, respectively). Estes et al. (1975) did not report a correlation for teacher judgment and the ERAS. It is interesting to note that the TRRS teacher estimates correlate somewhat lower with the ERAS indicating that children's estimates of attitude toward reading may be more valid. Dulin and Chester (1974; Chester & Dulin, 1977) suggest that children's responses may provide the most reliable criterion in attitude assessment. In validating an instrument using 36 photographs to assess attitudes toward reading for kindergarten, grade one, and grade two children, Redelheim (1976) also found that children's scores on the attitude measure were more closely associated with their self-ratings of attitude than with ratings made by their teachers.

Data from previous studies also indicate that achievement in a subject area will correlate positively with attitude measures in that subject. Thus, correlations between reading achievement and reading attitude measures provide another indication of convergent validity. The ERAS correlates .33 with the MAT Vocabulary scores and .37 with the MAT Comprehension scores. Both correlations, although modest, are positive and significant ($p < .001$) and the two correlations, as might be expected, are similar in magnitude. These validity correlations are somewhat lower than the correlation of .50 reported by Estes et al. (1975) for the ERAS and the STEP test using older subjects. This suggests that attitude may become more important in relation to school subjects as children grow older. It would be interesting to see if this relationship holds with younger and older subjects in other content areas as well. No data were reported in the Dulin and Chester validity study on the correlation between reading attitude and achievement.

Conclusions

There is a need for good instruments which are capable of providing accurate estimates of children's attitudes toward reading. Once instruments are available, further research can be conducted to indicate how attitudes develop and how they can best be influenced in a positive direction. Given the state-of-the-art in affective

assessment, chances for success are currently greatest if a combination of measures are used including both reactive and nonreactive types. Reactive measures are those in which subjects are aware a measure is being taken and include questionnaires, interviews, repertory grid techniques, and inventories. The questionnaire or inventory is the most frequently reported reactive measure now being used in measuring reading attitude (Summers, 1977). Nonreactive measures are those in which subjects are not aware a measure is being taken and include observation checklists and unobtrusive archival records such as library circulation data and classroom indices. On the basis of the validity data from this study, it seems reasonable to conclude that the ERAS is useful in obtaining a global reactive assessment of school-oriented attitude toward reading in intermediate grades. The internal consistency and convergent validity findings also confirm results obtained in previous studies using older subjects. However, with a reported readability level of grade 6, users of the ERAS should be alert to the possibility that some poor readers may need help in responding to the scale. It should also be borne in mind that these conclusions are based on responses to a given instrument by the specific group of subjects surveyed in this study.

References

Alexander, J. E., & Fillmore, R. C. *Attitudes and reading.* Newark, Delaware: International Reading Association, 1976.

Anastasi, A. *Psychological testing.* New York: Macmillan, 1976.

Athey, I. Research in the affective domain. In H. Singer & R. B. Ruddell (Eds.), *Theoretical models and processes of reading.* Delaware: International Reading Association, 1976.

Chester, R. D., & Dulin, K. L. Three approaches to the measurement of secondary school student's attitudes toward books and reading. *Research in the Teaching of English,* 1977, *8,* 193-200.

Cronbach, L. J. Coefficient Alpha and the internal structure of tests. *Psychometrika,* 1951, *16,* 297-334.

Dulin, K. L., & Chester, R. D. A validation of the *Estes Attitude Scale. Journal of Reading,* 1974, *18,* 56-59.

Durost, W. N., Bixler, H. H., Wrightstone, J. W., Prescott, G. A., & Balow, I. H. *Metropolitan Achievement Tests: Teacher's handbook, teacher's directions, intermediate and advanced forms.* California: Harcourt, Brace Jovanovich, Inc., 1970-71.

Ebel, R. E. *Essentials of educational measurement.* Englewood Cliffs, New Jersey: Prentice Hall, 1972.

Estes, T. H. A scale to measurement attitude toward reading. *Journal of Reading,* 1971, *15,* 135-138.

Estes, T. H., & Johnstone, J. P. Assessing attitude toward reading: A validation study. In P. L. Nacke (Ed.), *Interaction: Research and practice for college-adult reading.* Twenty-third Yearbook of the National Reading Conference, 1974, 219-223.

Estes, T. E., Johnstone, J. P., & Richards, H. C. *Estes attitude scales.* Charlottesville, Virginia. Virginia Research Associates Ltd., 1975.

Ewing, J. M. Assessing attitudes in reading. Paper presented at the 2nd Annual Meeting of the Reading Association of Ireland, Dublin, Ireland, 1977 (ERIC Document Reproduction Service Order No. ED 162 250).

Gephart, W. S., Engle, R. B., & Marshall, F. J. *Evaluation in the affective domain.* Bloomington, Indiana: Phi Delta Kappa, 1978.

Guilford, J. P. *Fundamental statistics in psychology and education.* New York: McGraw-Hill, 1956.

Khan, S. B., & Weiss, J. The teaching of affective responses. In R. M. W. Travers (Ed.), *Second handbook of research on teaching.* Chicago: Rand McNally & Co., 1973.

Krathwohl, D. R., Bloom, B. S., & Masia, B. B. *Taxonomy of educational objectives: Handbook II: Affective domain.* New York: David McKay, 1964.

Nelson, L. R. *Guide to LERTAP use and interpretation.* Dunedin, New Zealand: University of Otago, 1974.

Redelheim, P. S. A multidimensional test of reading attitude for children. *The Reading Teacher,* 1976, *29,* 181-186.

Reed, K. Reading instruction: The affective domain "Measuring affective growth in reading." Paper presented at the 7th Annual Meeting of the International Reading Association World Congress on Reading, Hamburg, West Germany, 1978 (ERIC Document Reproduction Service No. ED 160 990).

Shaw, M. E., & Wright, J. M. *Scales for the measurement of attitudes.* New York: McGraw-Hill, 1967.

Summers, E. G. Instruments for assessing reading attitudes: A review of research and bibliography. *Journal of Reading Behavior,* 1977, IX, 137-165.

GLEN G. FOSTER
Palm Beach County School System
BOB ALGOZZINE
University of Florida
and
RONALD C. EAVES
Auburn University

Predicting the Educational Bias Effect

Bias effect research has been conducted over the past 15 years, with many investigations relating to the educational system. The authors note that isolated investigations may be looked at together to build a logical argument that biasing conditions indigenous to the public school delivery system do influence teacher perceptions, thereby influencing actual child behaviour. Laboratory techniques for investigating the educational bias effect have progressed to the point where investigators may begin to study the variability in a group's susceptibility to this effect. This paper represents one such effort and documents a significant relationship between teachers' acceptance of diverse behaviour and their resistance to the biasing influence of a deviancy label.

The term educational bias effect refers to the phenomenon whereby a teacher's expectations for a child's behaviour may actually become a determinant of the child's behaviour. This area of research has its antecedents in the experimenter bias effect isolated by Rosenthal and Fode (1963). These researchers noted that experimenters with preconceived notions of the "proper" outcomes of an experiment inadvertently influenced human subjects to perform in ways consistent

---

Glen G. Foster received the Ph.D. in School Psychology from The Pennsylvania State University in 1975; he is currently in private practice but maintains an avid interest in the problems of labeling children.

Bob Algozzine, Ph.D. (Penn State), is an Associate Professor of Special Education at the University of Florida; his interests include ecological interpretations of exceptionality and issues in the assessment of children. He is currently involved in the University of Minnesota's Institute for Research on Learning Disabilities where he is a Research Associate studying bias in assessment.

Ronald C. Eaves is an Associate Professor of Special Education at Auburn University; he received his doctoral degree from the University of Georgia. His interests include improving assessment practices and training teachers of emotionally disturbed children.

with the experimenter's bias. Although attempts to replicate Rosenthal and Fode's study have been to some extent unsuccessful (Barber, Calverley, Forgione, McPeake, Chaves, & Bowen, 1969), the term experimenter bias effect was coined and researchers began to investigate the phenomenon's influence in social situations outside the research laboratory (i.e., school classrooms).

Palardy (1969) demonstrated that existing teacher expectancies were related to the academic performance of first grade children. His study differentiated between teachers who believe boys and girls learn to read with equal facility and teachers who held to the notion that, relative to girls, first grade boys are handicapped in learning the basic reading skills. Palardy concluded:

> When first grade teachers reported that they believed that boys are far less successful than girls in learning to read, the boy pupils of these teachers did achieve less well on a standardized reading test than a comparable group of boy pupils whose teachers reported that they believed that boys are as successful as girls in learning to read. (p. 374)

Researchers next investigated variables indigenous to the educational system which have the potential to influence a teacher's expectations for a child's success in school. Jones (1970) showed that perceived sociocultural level influences teachers' expectations of a child's self-perceptions. Cahen (1966) demonstrated that IQ and reading group assignment were sufficient as biasing factors to alter student teachers' grading of test booklets. Ross (1974) found that physical attractiveness is a determinant of teacher expectations. When exposed to a standardized and nebulous psychological report, experienced classroom teachers more often felt attractive children belonged in regular education classes while unattractive children would best be served through special education resources. A variety of other factors have been shown to be influential in biasing teachers' expectations for their students. Brophy and Good (1974) have provided an excellent review of this literature; they indicated that positive and negative "halo effects" operate in "teachers' ratings of students just as in any interpersonal ratings" (p. 135). Within the area of special education, research has consistently demonstrated that the series of medical model deviancy labels utilized by the public schools generate lowered expectations for child performance. Foster and Ysseldyke (1976) investigated the relationships between teacher expectations and the labels of mentally retarded, emotionally disturbed and learning disabled. Using a laboratory technique, it was shown that each of these labels lowered expectations and that this halo effect was sufficient in strength to cause experienced teachers to misinterpret videotaped child performance. Further investigations have demonstrated that this phenomenon operated even when the teachers were cautioned to be objective in their ratings of the videotaped behaviour (Foster & Salvia, 1977). A variety of other studies have shown that teachers react negatively to special education labels as expectancy generating stimuli (Algozzine & Mercer, in press).

Teachers' relative acceptance of diverse behaviour has been the focus of recent research. Algozzine (1977) has shown that teachers find disturbing behaviours differentially bothersome while Herr, Algozzine and Eaves (1976) have shown reactions to disturbing behaviour may be modified with training. Algozzine, Mercer and Countermine (1977) demonstrated that acceptance of a child's behaviour was a function of whether that behaviour was or was not expected based on the categorical label assigned the child. Thus variance exists in the degree to which

teachers find certain behaviours disturbing and the above investigations have refined a technique to measure this variability.

This investigation represents research within a fourth generation of bias effect studies. The first generation of work isolated the effect in a non-educational setting, the second generation demonstrated that the effect influences actual child behaviour in an educational setting and the third generation has demonstrated that the educational bias effect may be elicited by a variety of variables present in our educational system. The fourth generation of research, as represented here, has investigated teacher variability as regards susceptibility to the bias effect. It is anticipated that such work will eventually lead to the prediction of the effect and suggestions to minimize the detrimental results that bias has for individual children.

Method

Subjects

Subjects consisted of 28 regular education classroom teachers, 25 female and three male. The teachers were attending a three-day continuing education workshop offered through a large southern university. Mean age of the subjects was 31.2 years and mean years of experience was 5.8 years. The subjects were about evenly distributed over the public school grade range with 29% teaching K-3, 25% teaching 4-6, 25% teaching 7-9, and 21% teaching grades 10-12. Nineteen subjects held Bachelor's degrees while 9 held Master's degrees.

Materials

Disturbing Behavior Checklist (DBC). The Disturbing Behavior Checklist (DBC) is a factor analytically derived scale of behaviours which have been shown to be differentially bothersome (i.e., disturbing) to teachers. It contains 55 items, most of which can be arranged into four mutually exclusive factors. Sets of behaviours which are bothersome because they are immature, socially defiant, physically disturbing, and socially delinquent have been identified (Algozzine, 1977). The validity of the scale has been evaluated using logical, correlational, and experimental procedures; results of these studies have been reported by Algozzine, Barnette, and Danielson (1977). The reliability of the scale and subscales (i.e., coefficient *alpha*) has been reported to range from .62 to .93 (Algozzine, 1977) and was considered adequate for use of the DBC in this research.

In responding to the checklist, subjects are requested to rate the relative "disturbingness" of the 55 items of which it is comprised. The responses are placed on a 1-5 Likert-type scale with 1 = not very disturbing and 5 = very disturbing. Thus high scores are indicative of a respondent finding items (behaviours) to be relatively disturbing. Respondents are instructed to rate each item with regard to how disturbing it would be in working with children.

Videotape. Bias was measured by teacher reactions to a 12-minute colour videotape of a normal fourth grade male. The tape showed the child taking subtests of the Wide Range Achievement Test, the Peabody Individual Achievement Test, several psychomotor exercises and engaging in a free play period. The child was selected from a pool of over 400 children on the basis of normalcy in the areas of intellectual, academic and social development, and physical appearance (Foster, Ysseldyke & Reese, 1975).

F-Y Checklist. Reactions to the videotape were measured by responses on a 23-item referral instrument developed for use with this tape. Subjects were asked to rate the child's behaviour in various areas including academic skills, perceptual/motor development, activity level and personal/social adjustment. Ratings were made along a 106 mm continuum divided into five levels from "superior" to "far below average." Distance along this line constituted the score for each item, with the average distance for all items constituting the total score for the measure. Low scores are indicative of positive ratings and scores above the midpoint (i.e., 53) are considered below average expectations. The total score for the F-Y Checklist consists of the average for all items completed; this score was seen as an appropriate measure since previous research had indicated that subscores within the scale (i.e., for academics or activity level) were not significantly different from this total (Foster, 1976).

Procedure

Data for this investigation were collected in two separate settings. Subjects were asked to complete the DBC during the first session and to rate the videotaped behaviour during the second session held two days later. Separate researchers conducted the two sessions and the subjects were not told that the two data collection points were related.

During the bias effect portion of the study, subjects were given the F-Y Checklist and told that this instrument would be used to provide responses to a videotaped child. The tape was accurately explained to the subjects except they were not told that the child was normal. Instead, the subjects were told that the child had been diagnosed by a clinical team as being retarded and this label represented the biasing information given the teacher. The tape was shown, F-Y Checklists completed and collected, and the subjects were debriefed concerning the true nature of the research.

Results

The F-Y Checklist yielded one score for each teacher, representing an overall mean for the responses to the 23 items of the instrument. The Disturbing Behavior Checklist yielded five scores, corresponding to means for the four DBC factors and an overall mean. Group means and standard deviations for DBC and F-Y Checklist responses are presented in Table 1. The descriptive data suggest that the teachers rated the child's behaviour as below average (i.e., above midpoint of 53) and found the behaviour of the DBC to be moderately disturbing.

A set of six dependent *t* tests was utilized to determine if the observed differences present in teachers' ratings of the disturbingness of the DBC factors were statistically significant. The results of these tests are posted in Table 2; nonsignificant differences are underlined. The behaviours comprising the socially defiant factor of the DBC (i.e., destructiveness, disobedience, negativism) were seen as more bothersome than those of the other three factors. The results are consistent with those of other investigations (Algozzine, 1976, 1977).

To determine the relationship between teachers' ratings on the DBC and their responses to the F-Y Checklist, correlation coefficients and proportion of variance estimates were computed (Ferguson, 1971); the correlations are presented in Table 3. A significant relationship was found between the F-Y Checklist responses and

TABLE 1
MEANS AND STANDARD DEVIATIONS FOR SUBJECTS' RESPONSES TO DBC AND F-Y QUESTIONNAIRE

F-Y	Overall DBC Score	Social Immaturity	Socially Defiant	Physically Disturbing	Socially Delinquent
62.79	3.5	3.2	3.9	3.5	3.3
7.93	0.4	0.4	0.6	0.9	0.7

TABLE 2
RESULTS OF *T* TESTS COMPARING RESPONSES TO FOUR DBC FACTORS

Socially Defiant	Physically Disturbing	Socially Delinquent	Social Immaturity
3.9	3.5	3.3	3.2

TABLE 3
CORRELATION COEFFICIENTS BETWEEN F-Y QUESTIONNAIRE AND DBC FACTORS

	Overall DBC Score	Social Immaturity	Socially Defiant	Physically Disturbing	Socially Delinquent
F-Y Checklist	.44*	.38*	.18	.13	.30

* $p < .05$

the overall DBC ratings, accounting for about 20% of the variance. Only the socially immature subscale of the DBC yielded a significant relationship with the F-Y Checklist, accounting for about 14% of the variance.

Discussion

The results of this administration of the F-Y Checklist are consistent with previous data collected under similar conditions. Previous research (Foster & Ysseldyke, 1976; Foster, Ysseldyke & Reese, 1975; Foster & Salvia, 1977) has demonstrated that F-Y Checklist ratings, when beyond the midpoint rating of 53, are indicative of the bias generated by the described research technique and, in fact, constitute a halo effect measurement. The purpose of this investigation was to find correlates to the bias effect, not to duplicate previous research by demonstrating that the technique results in establishment of a bias. Readers interested in the establishment and measurement of the bias effect are referred to these previous studies.

Teachers' responses to the socially defiant factor of the DBC were significantly different from their responses to any of the other factors. These results are similar

to those of other investigations (Algozzine, 1976; Curran, 1977; Mooney & Algozzine, 1978) and lend support for the validity of this administration of the instrument.

The significant correlation between bias effect ratings (F-Y Checklist) and DBC ratings suggests that those teachers who found the DBC behaviours to be less disturbing were also less susceptible to the bias effect. The DBC factor of social immaturity was the best single factor predictor of checklist scores and was the only factor to obtain significance as an isolated predictor. This makes intuitive sense as it is logical that the stereotype for retarded children would contain elements relating to immaturity. Teachers who found immature behaviour more disturbing were also most susceptible to the biasing influence of the "mentally retarded" label.

Research links exist which will allow us to logically deduce that a teacher's susceptibility to bias will influence the academic performance of certain children. This investigation has established that a relationship exists between a teacher's reported acceptance of diverse behaviour and his/her susceptibility to the bias effect. It seems to follow that the degree of acceptance may also relate to academic performance of at least some children.

This research has also demonstrated that a laboratory technique may successfully be used to investigate variance in susceptibility to the educational bias effect. It is felt that this report of the relationship between susceptibility to bias and non-acceptance of behaviour should serve as an impetus for further research, rather than as the basis for decisions within the field.

Anticipating the day when the bias effect is openly admitted to operate in ways harmful to children and the relationship between bias and non-acceptance is verified, we offer the following suggestions to reduce the negative effects:

1. School system personnel may wish to reduce the number of negatively biasing variables indigenous to the classroom setting (e.g., deviancy labels, rampant use of mass administered intelligence and achievement tests).
2. We may wish to re-examine our present emphasis on homogeneous grouping patterns. Smaller class sizes with expanded ability ranges would allow for more individualization of instruction and less artificiality in our expectations of what children should or should not be like.
3. Preservice and inservice training in broadened acceptance of behaviour could be instituted within universities and school systems. Ongoing emphasis toward acceptance would be more effective than any short term efforts to alter teachers' expectations.
4. School system personnel may wish to make use of these findings in their selection process for new teachers. A teacher's tolerance toward diverse behaviour could be assessed and could be considering along with GPA, GRE scores, and other personal information.
5. Sophisticated matching of teacher and pupil may one day become an empirical and administrative possibility. The data presented here would indicate that teachers who find many behaviours to be disturbing may be less effective in working with certain children as they will be more negatively influenced by the biasing labels these children carry.

References

Algozzine, B. The emotionally disturbed child: Disturbed or disturbing? *Journal of Abnormal Child Psychology,* 1977, *5,* 205-211.

Algozzine, B. The disturbing child: What you see is what you get? *Alberta Journal of Educational Research,* 1976, *22,* 330-333.

Algozzine, B., Barnette, J. J., Danielson, L. The disturbing child: A validation study. Paper presented at the Annual Conference of the Council for Exceptional Children, Atlanta, Ga., April, 1977.

Algozzine, B., & Mercer, C. D. Labels and expectancies for handicapped children and youth. In D. A. Sabatino & L. Mann, *Fourth review of special education,* in press.

Algozzine, B., Mercer, D., & Countermine, T. The effects of labels and behavior on teacher expectations. *Exceptional Children,* 1977, *44,* (2), 131-132.

Barber, T., Calverly, O., Forgione, A., McPeake, J., Chaves, J., & Bowen, B. Five attempts to replicate the experimenter bias effect. *Journal of Consulting and Clinical Psychology,* 1969, *33,* 1-6.

Brophy, J., & Good, T. *Teacher-student relationships: Causes and consequences.* New York: Holt, Rinehart & Winston, 1974.

Cahen, L. Experimental manipulation of bias in teachers; scoring of subjective tests. Paper presented at the meeting of the American Psychological Association, New York, September, 1966.

Curran, T. Mainstreaming attitudes as a function of behavioural expectations. Unpublished doctoral dissertation, The Pennsylvania State University, 1977.

Ferguson, G. *Statistical analysis in psychology and education.* New York: McGraw-Hill, 1971.

Foster, G. Expectancy and halo effects as a result of artificially induced teacher bias. Unpublished doctoral dissertation. The Pennsylvania State University, 1976.

Foster, G., & Salvia, J. Teacher response to the label of learning disabled as a function of demand characteristics. *Exceptional Children,* 1977, *43,* 533-534.

Foster, G., & Ysseldyke, J. Expectancy and halo effects as a result of artificially induced teacher bias. *Contemporary Educational Psychology,* 1976, *1,* 37-45.

Foster, G., Ysseldyke, J., & Reese, J. I wouldn't have seen it if I hadn't believed it. *Exceptional Children.* 1975, *41,* 469-474.

Herr, D., Algozzine, R., & Eaves, R. Modification of biases held by teacher trainees toward the disturbingness of behavior. *Journal of Educational Research,* 1976, *69,* 261-264.

Jones, R. Labeling black college students culturally disadvantaged: A search for behavioral correlates. Paper presented at the meeting of the California Educational Research Association, April, 1970.

Maurer, A. Whatever happened to witches. *Journal of School Psychology,* 1972, *19,* 107-110.

Mooney, C., & Algozzine, B. A comparison of the disturbingness of behaviors related to learning disability and emotional disturbance. *Journal of Abnormal Child Psychology,* 1978, *6,* 401-406.

Palardy, J. What teachers believe, what children achieve. *Elementary School Journal,* 1969, *69,* 370-374.

Rosenthal, R., & Fode, K. Psychology of the scientist: Three experiments in experimenter bias. *Psychological Reports,* 1963, *12,* 491-511.

Ross, M. Attractiveness as a biasing factor in teacher judgments regarding special class placement and related factors. Unpublished doctoral dissertation, The Pennsylvania State University, 1974.

PERSPECTIVES

Research Traditions in the History of Education: A Study of Goals and Their Emergence

Henry W. Hodysh
The University of Alberta

Eighteen years have elapsed since Thomas Kuhn first published his study of the paradigm and its function in the history of science (1970a, first published 1962). Although Kuhn's interpretation of the paradigm as an organizing principle of research has stimulated extensive discussion in many fields of inquiry (Kuhn, 1970b, 1970c), it has generated little activity among historians of education[1] who, it would seem, have remained somewhat distant from his concerns. A number of reasons might account for this inattention, though perhaps it would not be inappropriate to speculate that many historians of education have viewed Kuhn's methodological interests as subordinate to the more established research problems in the field.

Some progress in correcting this oversight might be achieved by exploring the implications of Kuhn's thesis for the way in which educational history has emerged as a field of study in the European setting. In determining these implications, the effect of social and educational forces external as well as internal to the community of historians itself might be taken into account (Kuhn, 1970a, p. x). Such an approach would be of assistance in complementing the genetic or linear perspective so frequently adopted towards the study of the past (Hodysh, 1977). Attention might be directed to the concept of the paradigm as a first step in this endeavour.

For Kuhn, the operation of the paradigm as an organizing principle in the development of knowledge is exemplified in the following observation:

At the start a new candidate for paradigm may have few supporters, and on occasions the

---

Dr. Hodysh is a Professor in the Department of Educational Foundations at The University of Alberta. He has served as Assistant Dean in the Faculty of Education and is currently Editor of the AJER. Historiography, teacher education, and curriculum are his special research interests.

> supporters' motives may be suspect. Nevertheless, if they are competent, they will improve it, explore its possibilities, and show what it would be like to belong to the community guided by it. And as that goes on, if the paradigm is one destined to win its fight, the numbers and strength of the persuasive arguments in its favor will increase. More scientists will then be converted, and the exploration of the new paradigm will go on. (1970a, p. 159)

This assertion entails two interdependent meanings of paradigm, both of which are integral to Kuhn's position. On the one hand, "paradigm" represents the "disciplinary matrix" or "the entire constellation of beliefs, values, techniques, and so on shared by members of a given community." On the other, it "denotes one sort of element in that constellation, the concrete puzzle-solutions which, employed as models or exemplars, can replace explicit rules" as a basis for solving the puzzles of normal science (1970a, p. 175). The paradigm as puzzle-solution, therefore, is contained within and dependent upon the interaction of elements in the disciplinary matrix. In sum, the paradigm as disciplinary matrix plays a foundational role in the development of knowledge in the scholarly community. It is the ground or basis of the essential characteristics and components which define the community's research traditions (Kuhn, 1977, p. 297).

If the paradigm can be employed as a basis for the study of research traditions in the scientific community, what, then, are its implications for the analysis of research traditions in the history of education? Insight into this question might be gained by focusing more closely on the way in which historians of education perceive the development of goals in their field of study. According to Kuhn, members of a professional community see themselves "and are seen by others as the men uniquely responsible for the pursuit of shared goals" (1970a, p. 177). It is a knowledge of these goals as shared commitments that becomes critical to our understanding of the disciplinary matrix, for such knowledge helps to establish an essential component of the community's research traditions.

The problem might be more clearly defined in relation to the following question: In what way are the goals and commitments of educational historians reflected in their historical research? The question is critical to our study for it directs us to the concerns of the historian and, as such, serves as an end or focal point for the research activities of the community. Its significance in the study of research traditions might be clarified through an examination of the community's professional literature (Kuhn, 1977, pp. 35ff) — a task which recognizes that a field of study is defined not only by elements integral to its disciplinary matrix but, as in the case of intellectual history, by forces extant and operative in the broader cultural-historical milieu (Kuhn, 1977, p. xv).

By way of illustration, then, this study explores developments in the European context with particular reference to Polish educational historiography during the first quarter of the twentieth century.[2] The setting is of special interest, for it was during that time — after some 125 years of partition and occupation by Austria, Prussia, and Russia — that Poland achieved independence and embarked upon a program of intellectual development which extended into many areas of life (*Nauka Polska*, Vol. I, 1918; Nawroczynski, 1938, pp. 173-193), including the study of history and history of education (Morawski, 1922, p. 14; Konopczynski, 1925, p. 4). In many ways it was a period of intellectual renewal (Perkowska, 1975) in which historians refurbished their links with a rich and colourful past and, at the same time, looked ahead to the future (Dyboski, 1933-34, p. 318). Moreover, it was a time

when the history of education contributed in a special way to the nation's cultural identity (Barycz, 1949, 1967; Hulewicz, 1974).

The study, citing selected developments in Polish educational historiography, focuses specific attention on the professional goals and commitments of the community of historians as factors central to the emergence of its research traditions. It endeavours to determine the nature of these goals through a descriptive analysis of the community's professional literature. Within this general framework, both the possibilities and limitations of Kuhn's position are considered without necessarily accepting its particulars.

The Emergence of Educational History

The development of Polish historiography during the years preceding independence in 1918 is of particular interest to the growth of educational history as a field of study. Although the partitioning powers tended to neglect the study of Polish history, using it as a means to advance their own interests (Karbowiak, 1900, p. 8), Polish historians themselves continued their historical research and contributed a significant body of literature to the nature and purpose of the field (Bujak, 1930; Dyboski, 1933-34). The efforts of historians to describe and explain the development of the nation's past resulted in the formation of the Cracow and Warsaw Schools of Historical Thought (Dutkiewicz & Sreniowska, 1959). Whereas the Cracow School perceived the setbacks and defeats that led to partition to be the result of internal problems, the Warsaw School attributed the nation's misfortunes to external developments and political events beyond her control (Kot, 1941, p. 32; Serejski & Grabski, 1974, pp. 5-34). These differing perspectives influenced much of the discussion regarding the function of history during the latter part of the nineteenth and early part of the twentieth century.

Within this setting, the history of education was studied as part of the nation's cultural history. As a special interest, it was represented in the writings of Szujski (1878), Windakiewicz (1890), Morawski (1900), and others (Hulewicz, 1960) and, in terms of its formative development, could point to a tradition that reached as far back as the eighteenth century (Lempicki, 1925, pp. 1-2).

By the early 1900's, scholars at the Jagiellonian University in Cracow and the University of Lwow suggested that the history of education as a study required greater attention and development. At the Third Congress of Polish Historians, for example, Karbowiak advocated a reconstruction of archival and monographic materials in the field, proposing that a closer look at Polish educational history would create opportunities not only for an examination of problems particular to education but for awakening the younger generation to the "spirit of the nation" (1900, p. 7). Karbowiak argued that these objectives might best be achieved by establishing academic chairs in Pedagogy and the History of Education at Polish universities. His suggestion was echoed later by Kukulski who called for a program in the History of National Education at the Jagiellonian in Cracow. Alluding briefly to the historical debates of the nineteenth century, Kukulski also suggested the need for new directions in the interpretation of Poland's educational past (1914). It would seem, therefore, that an intellectual climate conducive to educational history as a specialization within the broader field of cultural history was already in the making before the outbreak of World War One.

The achievement of Polish independence following the war opened a new chapter in the history of education as a field of study. In Cracow on November 5,

1919, the Minister of Religious Denominations and Public Education, Lukasiewicz, charged a commission under the direction of Morawski, President of the Polish Academy of Sciences and Letters "to fill a great vacuum in the area of pedagogical thought and practice."

> In relation to this aim, the Ministry has called to life a Commission of History and Pedagogy to research the history of education and schools in Poland. The Commission has prepared a very broad plan of educational activity and has hopes of a series of studies and monographs in the history of education and pedagogical thought. (column CL/42)

During its first years of operation, the Commission gained vigorous support from historians and educators. In a newsletter from the Lwow section of the Commission, Lempicki commented upon the importance of its activities to the nation.

> . . . with the birth of our homeland . . . it is necessary to build a deep feeling of nationhood and its system of education: we must look to our educational past which will serve as a bond to the present system of education . . . (1922, p. 2)

The Commission contributed to a renewed interest in the history of pedagogical thought (Barycz, 1949, pp. 30-37) and led to the publication of a variety of historical investigations, including among them studies of Konarsky, Piramowicz, Kollataj, Sniadecki, the history of parish schools, and the famous Commission of National Education of 1773 (Bobkowska, 1937, p. 143; Ministerstwa . . . Publicznego, 1921, pp. 167ff), Europe's first ministry of education (Dyboski, 1933, p. 270).

Organization and Development

A key figure in the organization and development of research in educational history during the first decade of independence was Professor Kot from the Department of Cultural History at the University of Cracow. In 1913, Kot had visited archives throughout Europe in order to determine the place of Polish scholarship in the context of the European historical community (*Sprawozdanie . . . Akademii Umiejetnosci,* 1914, p. 6). Later, in his role as Secretary to the Pedagogical Commission, he coordinated a cadre of experts from Cracow, Lwow, Warsaw, Poznan, and other centres in organizing and developing archival materials. As author and editor of many scholarly publications, Kot turned his attention to a number of areas for research, most notably the history of education, the Reformation of the sixteenth and seventeenth centuries in Poland, and the relation of Polish thought to the wider European historical scene (Lempicki & Hartleb, 1937, pp. 11-12). His participation in learned societies provided visibility to the history of education and placed it among the developing interests of historical and educational scholarship in Poland (Tazbir, 1976, pp. 752-755). Bruckner, a prominent historian of the period, noted Kot's interest in the history of education as a stroke of good fortune for a "much neglected" field of inquiry (1910).

Among Kot's studies of the past was the *Historja wychowania (The History of Education),* a two-volume cultural history of pedagogical thought and practice that placed the development of Polish education in the mainstream of European thought (1924). The study, though directed to the interests of the Polish historical community, was international in scope. It acknowledged not only the role of recent

Polish scholarship in the history of education but also the contribution of such writers as Barth, Compayre, and Paulsen (Kot, 1924, pp. 20-21).

Kot perceived the history of education to be part of the history of culture. As a field of inquiry, it was expected to meet the rigorous standards applicable to all historical scholarship and, as such, to be studied objectively and "founded upon the fullest of documentary evidence" (1924, p. 11). Utilizing a perspective that encompassed both institutional and noninstitutional education as topics for investigation, Kot introduced references to economic, artistic and other cultural factors as important in establishing the dominant "pedagogical ideal" in any given period of time (Kot, 1924, p. 7). The *Historja* begins with a description and explanation of education in classical Greece and Rome. This is followed by educational developments as they occurred in the Medieval, Renaissance, and eventually, the Modern periods of history. In the Polish setting, attention is directed to Humanism, the educational reforms of Konarsky, the Commission of National Education of 1773, the Reformation and Protestant Schools, the Jesuits, education under three-power partition, and, finally, a brief commentary on educational developments in the post World War One period.

If the importance of an individual's contribution to his field is gauged in part by the assessments of his colleagues, then Kot's leadership in the development of Polish educational historiography deserves particular recognition. For example, a review of the *Zrodla do historja wychowania (Sources to the History of Education)* — a compilation and study of original source materials highlighting the development of pedagogical theory and practice — awarded high marks for his breadth and originality of scholarship (Bobkowska, 1930). It appears that Kot's efforts in the study and organization of educational history raised the field to a position of prominence, affecting in a significant way the historical research of his colleagues and students. (Barycz, 1967, pp. 202ff.).

In sum, a number of events critical to the formation of goals held by members of the community of historians occurred in the first quarter of the twentieth century. The community revitalized the study of Poland's educational past and, in the process, added a dimension to cultural history, the impact of which extended beyond the disciplinary matrix to the entire fabric of Polish intellectual life.

Discussion

It would be quite inappropriate to generalize on the basis of an exploratory study. Nevertheless, our inquiry lends itself to making a number of observations concerning both the possibilities and limitations of Kuhn's notion of the paradigm as disciplinary matrix — observations which centre on three categories or types of issues pertaining to the emergence and perception of goals in the field of educational history. The first category points to the influence of the socio-historical milieu on the function of educational history; the second relates to the nature of historical inquiry *qua* historical inquiry; and the third reflects some of the problems or limitations connected to Kuhn's notion of the paradigm that bear tangentially on the Polish data but which, nonetheless, are significant to further research.

The study suggests that the perception of goals held by members of the historical community may be shaped by developments immediate not only to the community itself but, as indicated by Kuhn, to a wider range of factors in society in

general. The achievement of Polish independence in 1918 appears to have been one such development. Its impact was made evident in a variety of contexts, including the creation of the Pedagogical Commission of 1919. To cite but one example, Lempicki's comment on the function and activity of the Commission reflects an interest in educational as well as historical questions confronting the newly created state. Such questions, therefore, were not the private preserve of historians, as they were discussed frequently by the scholarly community as a whole. Moreover, this does not imply that all historians of education were equally influenced by events in society or that they tended to interpret or view those events from an identical perspective. Rather, it is suggested that the professional activities of the historical community, including the selection of topics considered to be worthy of investigation, were affected to some extent by concerns common to all. However, the precise nature of this interaction between the community of historians and society as a whole is difficult to determine. It is a problem which in its own right necessitates further investigation.

In relation to the second issue, attention was directed to the nature of goals and commitments in terms of both the data of educational history and the framework in which the data were enclosed. For example, it would appear that Kot's selection and interpretation of historical events were accompanied by an awareness of conceptual questions addressed by historians of education. This awareness, as evidenced by Kot's reference to the nature of history of education and the contribution of his colleagues and predecessors to both the theory and practice of the study, served as a point of departure in the *Historja* for advancing two general kinds of historical inquiry. The first was a recognition of specialization and a concern for developing an in-depth body of knowledge based on monograph type studies in order to keep abreast of the ever-increasing quantity of historical data and research. The second was a recognition of the need to provide a synthesis of Polish history that would serve not only as an overview of the educational past but as a means of identifying and clarifying the course of the history of a nation which for more than a century had broken continuity. Kot's concern with both areas of inquiry in the history of education placed him in the vanguard of historical research in the field as it developed in the early decades of the century.

In the Polish setting, therefore, the history of education as a field of study was influenced not only by the interaction of elements within the disciplinary matrix itself, but also by events and conditions in the wider society not the least of which were the expectation and achievement of independence of 1918. On the one hand, the coming of independence challenged the historical community to study more closely the social, educational, and other issues confronting the newborn state, and, on the other, it encouraged the community to develop anew the tradition of research which was to extend beyond the historical questions of the nineteenth and into the years of the twentieth century.

It must be noted that our study of Kuhn's paradigm is illustrative only and that many historical developments have not been considered. How, for example, might Kuhn's analysis be employed to account for the apparent emergence of conflicting goals and commitments within the community? Additional study of what might be identified as a broader range of events, along with an analysis of the ideas of other historians and educators who influenced Polish educational historiography, would add to our understanding of research traditions.

In relation to the third category, the matter of limitations, it should be

reiterated that our inquiry has merely endeavoured to identify methodological possibilities and to explore a tentative rationale which historians of education have tended to overlook. The study has not defended a methodological alternative and, therefore, would be remiss if, however briefly, it did not take into account apparent limitations which the analysis has suggested.

The more obvious of these limitations pertain to the description of goals and their representation in the professional literature of the historical community. For example, the evidence by which goals and commitments are determined may not be readily apparent in an analysis of the professional literature of the community, necessitating a closer look at a broader spectrum of community involvements. Moreover, it might be necessary to explore community activities and societal conditions which are peripheral to the community itself, but which through their indirect influence bear upon the emergence and formulation of research traditions. A study of the informal literature as well as an analysis of the role and influence of other scholarly communities may hold insights into the nature and function of the historical community and its research traditions.

Beyond these rather specific problems are others no less significant, but more general in nature. For example, do educational historians comprise an extended group, numbering among its constituents not only academic historians but scholars from the field of education? Put differently, what criteria might be employed to define membership in the community of educational historians? Further, in what way do programs of teacher training and education affect the goals of the community and the determination of its research traditions?

Implications

Interestingly, the implications of Kuhn's position for the study of educational history are both particular and general. On the one hand, Kuhn's argument provides some direction to historians and, quite possibly, to others in related fields by suggesting specific focal points for investigation and further inquiry; on the other, it brings to attention the possibility of integrating some of Kuhn's ideas on the emergence of research traditions with existing methodological frameworks, a possibility which would seem to reflect a certain breadth and flexibility inherent in much of Kuhn's position.

The inquiry into the paradigm and its possibilities and limitations has been heuristic in purpose. No attempt has been made to draw definitive conclusions. Moreover, there are many aspects of Kuhn's position that have not been examined, such as the notion of the preliminary or pre-paradigm stage of a discipline's development. Shortcomings notwithstanding, the study has raised tentative though promising hypotheses — hypotheses which upon clarification and development might lead not only to a more critical analysis of Kuhn's argument but, perhaps, to a better understanding of educational history, its goals and research traditions.

Notes

1. The terms "history of education" and "educational history" are used interchangeably in this paper.
2. Diacritical marks found in Polish will be omitted due to complications in printing.

The writer gratefully acknowledges The Canada Council for providing a Research Grant used in the preparation of this study.

Sincere appreciation is also extended to members of The Polish Academy of Sciences in Warsaw, History of Education; The Pedagogical Institute, University of Warsaw; The Pedagogical Institute, Jagiellonian University of Cracow; and The Polish National Library in Warsaw for their assistance in obtaining Polish historical materials.

References

Barycz, H. Historia kultury, nauki, oswiaty i wychowania w Uniwersytecie Jagiellonskim. In S. Mikucki (Ed.), *Studia z dziejow Wydzialu Filozoficzno-Historycznego Uniwersytetu Jagiellonskiego.* Prace Historyczne. Zeszyt 16 (History of culture, learning, instruction and education at the Jagiellonian University. In *Studies from the history of the Faculty of Philosophy-History Jagiellonian University,* Historical Papers, Number 16). Cracow: 1967, 185-222.

Barycz, H. *Rozwoj historii oswiaty, wychowania i kultury w Polsce (Development of the history of learning, education and culture in Poland).* Cracow: 1949.

Bobkowska, W. Komisja do Dziejow Oswiaty i Szkolnictwa w Polsce, P.A.U., *Nauka Polska* (Commission to the History of Education and Schooling in Poland, Polish Academy of Sciences and Letters, *Polish Learning*) (Offprint). 1937, *22,* 141-146.

Bobkowska, W. Recenzje "Zrodla do historji wychowania." *Ruch Pedagogiczny* (Review of "Sources to the history of education." *Pedagogical Movement*), 1930, April, 169-172.

Bruckner, A. Recenzja. *Pamietnik Literacki* (Review. *Literary Memoirs*), 1910, *9,* 33.

Bujak, F. Kultura ludowa na tle kultury narodowej i powszechnej. *Kultura Wsi* (Folk culture against the background of national and universal culture. *Village Culture*). Warsaw: 1930. (Offprint)

Dutkiewicz, J., Sreniowska, K. *Zarys Historiografii Polskiej 1900-1939,* Vol. III *(Outline of Polish historiography 1900-1939).* Lodz: 1959.

Dyboski, R. Cultural problems of the new Poland. *The Slavonic Review,* 1933-34, *12,* 304-322.

Dyboski, R. *Poland.* London: 1933.

Hodysh, H. The Kuhnian paradigm and its implications for the historiography of curriculum change. *Paedagogica Historica,* 1977, *17,* 75-87.

Hulewicz, J. Dzialalnosc Wydawnicza Akademii Umiejetnosci w Krakowie w Zakresie Historii Literatury Polskiej oraz Historii oswiaty i szkolnictwa w Polsce w Latach 1873-1918. *Pamietnik Literacki* (Publishing activity of the Academy of Sciences and Letters in Cracow in the history of Polish literature as well as history of instruction and schooling in Poland 1873-1918. *Literary Memoirs),* 1960, *51,* 45-104.

Hulewicz, J. Historia wychowania i oswiaty. In E. Rostworowski et al. (Eds.), *Polska Akademia Umiejetnosci 1872-1952: Nauki humanistyczne i spoleczne* (History of education and instruction. In *Polish Academy of Sciences and Letters 1872-1952: Humanistic and Social Studies).* Cracow: 1974, 225-233.

Karbowiak, A. Stan badan na polu dziejow wychowania, szkola i nauk w Polsce. *Pamietnik III Zjazdu historykow polskich w Krakowie 1900* (State of research in the field of history of education, schooling and learning in Poland. *Memoirs of the III Congress of Polish Historians in Cracow 1900*), 1-11.

Konopczynski, W. Organizacja pracy historycznej. *IV Zjazd historykow polskich w Poznanju,* Sekcja II (Organization of historical works. *IV Congress of Polish Historians in Poznan,* Section II), 1925.

Kot, S. *Five centuries of Polish learning.* Oxford: 1941.

Kot, S. *Historja wychowania (History of education),* Vols. I and II (Lwow: 1924, second printing 1934).

Kot, S. *Zrodla do historji wychowania (Sources to the history of education).* Warsaw: 1929, Part I; 1930, Part II.

Kuhn, T. *The essential tension: Selected studies in scientific tradition and change.* Chicago: 1977.

Kuhn, T. *The structure of scientific revolutions.* Chicago: 1962, Second edition, enlarged, 1970. (a)

Kuhn, T. Logic of discovery or psychology of research? I. Lakatos & A. Musgrave (Eds.), *Criticism and the growth of knowledge.* Cambridge: 1970, 1-23. (b)

Kuhn, T. Reflections on my critics. In I. Lakatos & A. Musgrave (Eds.), *Criticism and the growth of knowledge.* Cambridge: 1970, 231-278. (c)

Kukulski, Z. Przedmowa. *Ktorzy Idziemy* (Introduction. *We Who Go Forward*), 1914, *2,* January-April, 1-5.

Kukulski, Z. W zwyczajnej Katedry Pedagogii i Dydaktyki, oraz Katedry Dziejow, Edukacy i Narodowej na Uniwersytecie Jagiellonskiem. *Ktorzy Idziemy* (Regarding a chair of pedagogy and teaching methods as well as a chair of history of national education at the Jagiellonian University. *We Who Go Forward*), 1913, *1*, October, 117-132.

Lempicki, S. Dzieje szkolnictwa i wychowania w Polsce. *IV Zjazd historykow polskich w Poznaniu* (History of the educational system and education in Poland. *IV Congress of Polish Historians in Poznan*), 1925, 1-18.

Lempicki, S. Komisja do badania dziejow wychowania i szkolnictwa w Polsce. *Wiadomosci z Dziejow Wychowania i Szkolnictwa w Polsce* (Commission for Research of the History of Education and the Educational System in Poland. *News of the History of Education and the Educational System in Poland*), 1922, *1*, 2.

Lempicki, S., & Hartleb, K. Historia kultury. *Kwartalnik Historyczny* (Cultural history. *Historical Quarterly*), 1937, *51*, 3-21. (Offprint)

Ministerstwa Wyznan Religijnych i Oswiecenia Publicznego. *Program Nauki w Panstwowych Seminarjach Nauczycielskich* (Ministry of Religious Denominations and Public Education. *Program of Studies in State Teachers' Colleges*). Warsaw: 1921.

Morawski, K. *Historya Uniwersytetu Jagiellonskiego (History of the Jagiellonian University)*, Vols. 1 and 2. Cracow: 1900.

Morawski, K. *Ku Czci Pracy i w Obronie Nauki (In reverence of work and in defence of learning)*. Cracow: 1922.

Nauka Polska (Polish Learning), 1918, *1*.

Nawroczynski, B. *Polska mysl pedagogiczna (Polish pedagogical thought). Lwow: 1938.*

Perkowska, U. *Kstaltowanie sie zespolu naukowego w Uniwersytecie Jagiellonskim, 1860-1920.* Monografie z dziejow nauki i techniki, XCVIII (*Formation of the academy of the Jagellonian University, 1860-1920.* Monograph of the history of learning and technology, XCVIII). Cracow: 1975.

Serejski, M., & Grabski, A. Michal Bobrzynski i jego Dzieje Polski w Zarysie. In M. Bobrzynski, *Dzieje Polski w zarysie* (Michal Bobrzynski and his Poland's history in outline. In M. Bobrzynski, *Poland's history in outline*). Warsaw: 1974, 5-34.

Sprawozdanie z Czynnosci i Posiedzen Akademii Umiejetnosci w Krakowie, Wydzial Filologiczny (Report of the Activities of the Academy of Sciences and Letters in Cracow, Faculty of Philology). Cracow: 1914.

Sprawozdanie stenograficzne ze 146 posedzenia Sejmu Ustawodawczego z dnia 11 maja 1920 r. (Stenographic report of the 146 sitting of the Sejm 11 May 1920), column CL/42.

Szujski, J. *Archiwum do dziejow literatury i oswiaty w Polsce (Archives to the history of literature and education in Poland)*. Cracow: 1878.

Tazbir, J. In memoriam: Stanislaw Kot. *Kwartalnik Historyczny* (In memoriam: Stanislaw Kot. *Historical Quarterly*), 1976, *83*, 752-755.

Windakiewicz, S. Znaczenie Szkol i uniwersytetow dla oswiaty i literatury Lacinskiej w Polsce. *II Zjazd historykow polskich w Lwowa* (The significance of schools and universities for Latin education and literature in Poland. *II Congress of Polish Historians in Lwow*), 1890.

ESSAY REVIEWS

TEACHING EDUCATIONAL STATISTICS IN THE COMPUTER AGE

Ross E. Traub
The Ontario Institute for Studies in Education

STATISTICAL METHODS IN EDUCATION AND PSYCHOLOGY. *By A. K. Kurtz and S. T. Mayo.* New York: Springer-Verlag, 1979, 538 pp. (ISBN 0-387-90265-1)

During the course of my lifetime, the computational power available to members of the general public, but especially to university students enrolled in statistics courses, has increased enormously. The only computational device available to me from 1943 to 1955, the years I attended school in a small Alberta town, was my father's analog computer, his slide-rule. It is true that in 1946, J. Presper Eckert and John W. Mauchly succeeded in building the first digital computer. But this invention did not affect Alberta students directly for more than ten years. As recently as 1958 when I enrolled in the introductory statistics course offered by Professor E. S. Keeping, all we had in the stat-lab were hand-cranked Munroe calculators. The computing resources then available to Alberta's educational researchers were not much better. During the summer of 1959, when I worked as a research assistant for Professors G. M. Dunlop and Donald Black, the principal computing aids to be found in the Faculty of Education were several electrically driven mechanical calculators. It wasn't until 1960, when I returned from a year of teaching school to work on a Master of Education degree, that I encountered the University of Alberta's first digital computer. Professor R. S. MacArthur had the

---

Ross Traub was born in Alberta and earned the B.Ed. and M.Ed. degrees from the University of Alberta. After studying at Princeton University for the Ph.D., Traub taught at McGill University for two years. Since 1966 he has been employed by OISE, teaching and conducting research and evaluation studies in the Department of Measurement, Evaluation and Computer Applications. He is currently Chairman of the Department.

prescience to insist I take a noncredit course in how to program it. Although the name of this computer has passed from memory, I do recall that it was difficult to program, primarily because instructions were fed into it by means of a continuous paper tape which had to be made with painstaking care. In view of developments since 1960, however, that noncredit course was one of the most profitable investments of time I ever made. Since 1960 the power of digital computers has increased almost exponentially, year by year. So too has their accessibility. In addition, the development of the LSI chip has meant that for a very few dollars anyone can acquire an electronic calculator small enough to be held in the hand, yet able to make the same computations faster and more quietly than the mechanical devices in use 20 years ago.

There is a point to be found in this bit of personal history. The world of computing has changed dramatically over the past twenty years. Instructors of statistics and the textbooks they use should be keeping pace; they should be making full and good pedagogical use of the computational power now available. Short-cut methods of "hand calculation" should interest historians of science, and no one else. Methods for grouping data to facilitate computation, Sheppard's correction which is needed to reduce the error in the standard deviation that is caused by grouping, and eight ways to obtain a square-root without the aid of a calculator are prime candidates for a museum display case. I can think of no good reason for consigning today's student to an academic purgatory from where safe passage is earned by making seemingly endless calculations, all to obtain only one or two statistics. Granted, there may be instructional value in a limited amount of "hand calculation," where the dimensions of the problem are the minimum required to reveal or to reinforce some concept. Beyond this point, however, students should be directed to the nearest available computer terminal, there to draw from storage data sets of realistic proportions and to investigate the characteristics of these sets by having the computer generate histograms, scatterplots, and all manner of descriptive and inferential statistics. To me, this approach to the teaching of statistics offers considerable promise: Understanding of concepts can be tested and modified through studies of varied and meaningful data sets; expertise can be acquired in handling data sets of realistic size; and, perhaps most important of all, the habit can be fostered of examining the match between assumption and data, of considering alternative statistics, and of trying different analytical models. (A good attempt in this direction, albeit one that is not directed specifically at budding educational researchers, is the *MINITAB Student Handbook* by Ryan, Joiner and Ryan, 1976.)

This sunny morning of the computer age is associated with the development of another significant emphasis in statistics, that of exploratory data analysis. The alternative and traditional emphasis is confirmatory analysis. In performing analyses of the latter type, the statistician uses data to test the validity of a prediction or hypothesis, something that was, or should have been, formulated through application of hypothetico-deductive logic well before the data were collected. In confirmatory analyses, statistics is used to test the match between empirical reality and substantive (as opposed to statistical) theory. On the other hand, the thrust of exploratory data analysis is very different. The operative metaphor is the process of detection as employed by a master sleuth. In exploratory analysis, an extant set of data is deemed interesting and worthy of study. It is examined closely in order to ascertain its salient features. Unsuspected

characteristics are noted. Clues are sought on which to base hypotheses or explanations. Ways by which to test hypotheses and explanations within the limits of the data are sought or invented. The statistician's objective is to provide a coherent description and interpretation of the data, not necessarily the only possible or even the best description and interpretation, just one that is internally consistent and plausible in the same way that a detective's description and interpretation of the data associated with a crime are internally consistent and plausible.

One feature of exploratory methods bears special mention. This is a concern the exploring statistician has for the effect of bad data. Exploratory analysts try to use methods that are robust, which is to say, they try to describe and interpret data using results that are insensitive to "outliers," i.e., bad data. Inasmuch as educational data in my experience are fraught with outliers, this emphasis is a welcome one indeed. (The testaments of exploratory data analysis have been written by John Tukey and Frederick Mosteller — c.f., Tukey, 1977; Mosteller & Tukey, 1977. An introductory text that incorporates the ideas of exploratory data analysis has been prepared by Canadian sociologists, B. H. Erickson and T. A. Nosanchuk, 1977).

Computational power is a great asset to the exploratory data analyst (although, as Tukey has shown, much can be done by hand). Computers facilitate the statistician's study of a data set. With a computer, a terminal, and a set of analytical routines that can be called upon in the interactive mode, it is easy for the statistician to explore quickly the characteristics of his data, and just as quickly to test out hypotheses, examine the effect of transformations, etc.

All of which brings us to the raison d'être for this review, the introductory statistics text by Albert K. Kurtz and Samuel T. Mayo. Briefly, this book is an anachronism. Kurtz and Mayo pay very little attention to the digital computer in their writing and they provide no basis at all for using it as a pedogogical aid. Moreover, the concept of exploratory data analysis and the topic of robust statistics are nowhere covered.

Having said this, I must admit that the book has some excellent features, and were this 1950 I'd probably recommend it highly. It's strengths are these: Good coverage of classical topics for a half-term (one-semester) course; plentiful and varied examples; a healthy respect for the standard error of a statistic — they report standard errors for all but one of the statistics they present; a breezy style of writing, well punctuated with humour; the inclusion of most possible interpretations of the statistics they present, some that are based on years of experience with educational and psychological data.

A chapter by chapter summary of the book's contents is as follows: Chapter 1 offers a rationale for statistics. It also contains a useful discussion of significant digits. Chapters 2 and 3 are devoted respectively to the topics of central tendency and variability. Not unexpectedly, in view of the lack of concern Kurtz and Mayo display for outliers, the mean and standard deviation are given favoured status. The fourth chapter is focussed on the normal curve. The relationship between this curve and the binomial frequency distribution is given appropriate emphasis. Statistical inference is the topic of Chapter 5. Many of the familiar concepts of confirmatory data analysis are discussed in this chapter — sampling distributions, the standard error of the mean, the t test, degrees of freedom, errors of inference (Types I and II), one- and two-tailed tests. A surprising omission, however, is the

concept of statistical power. Chapter 6 is focussed on percentiles and their calculation, Chapter 7 on the shape of a distribution, its skewness and kurtosis. Correlation and regression take centre stage in Chapters 8, 9 and 10. The presentation on Pearson's *r* includes illustrations of scattergrams for selected correlation coefficients and many examples of correlations from published studies. These examples suggest realistic expectations for the sign and magnitude of *r* for the kind of variables frequently encountered in educational and psychological research. This basis is augmented by an explicit set of statements from Kurtz and Mayo to be used in evaluating the *r* obtained in a study. Chapter 11 presents the chi-square statistic and illustrates its application to problems of contingency and to testing the fit of a theoretical distribution function to an observed frequency function. Several other nonparametric statistics are described in Chapter 12 — the sign test, the runs test, the median test, and the Mann-Whitney U test. This is followed by a chapter on one-way analysis of variance. In Chapter 14, a general framework is developed for testing differences between pairs of statistics. This framework is subsequently applied to means and standard deviations. The fifteenth and final chapter is entitled Reorientation. One purpose of this chapter is to confer "most important status" on six concepts: mean, standard deviation, Pearson's *r*, analysis of variance, statistical inference and normal distribution. Chapter 15 also briefs the reader about advanced courses and suggests uses for the statistics taught in this book.

In addition to the foregoing chapters, Kurtz and Mayo have provided 22 tables in an appendix. These include the usual entries for the normal, *t*, *F*, and chi-square distributions. Also, there are tables for use in testing the statistical significance of Pearson's *r*, point-biserial *r*, tetrachoric *r*, and the phi-coefficient. Consistent with their philosophy of ignoring the computational devices now available to students, Kurtz and Mayo have also tabled values of the function N(N–1) for the integers from 1 to 499, the products of all pairs of the integers from 1 to 34 with the integers from 1 to 49, and the square of N and the square roots of N and 10N for the integers from 1 to 999.

To conclude: If you and your students have access to a digital computer and if you, the instructor, wish to develop a respect for robust methods and interest and skill in exploratory data analysis, then this is *not* the statistics textbook you need. The only reason I can suggest for buying a copy is to reflect upon the strong opinions and recommendations of two experienced instructors.

References

Erickson, B. H., & Nosanchuck, T. A. *Understanding data.* Toronto: McGraw-Hill Ryerson, 1977.

Mosteller, F., & Tukey, J. W. *Data analysis and regression: A second course in statistics.* Reading, Mass.: Addison-Wesley, 1977.

Ryan, T. A., Joiner, B. L., & Ryan, B. F. *MINITAB student handbook.* North Scituate, Mass.: Duxbury Press, 1976.

Tukey, J. W. *Exploratory data analysis.* Reading, Mass.: Addison-Wesley, 1977.

RESEARCH NOTES

RELIABILITY OF THE ALBERTA ESSAY SCALES

The technique of grading essays through comparisons with a set of models has enjoyed at least intermittent popularity for several decades. The earliest scales (Hillegas, 1912; Willing, 1922; Van Wagenen, 1923; and Hudelson, 1923) remained in use until at least the 1930's. A scale produced by Educational Testing Services (1957) was listed in the Sixth Mental Measurement Yearbook (Buros, 1965, p. 592).

Nyberg and Nyberg (1977) developed two sets of essay models suited especially for grading compositions written by students at the grade 12 level in Alberta. The models were derived from essays written by students in English 30 in June, 1964. In that year, a random sample of 103 essays was drawn and graded by 21 raters for style and content, and by another 27 raters for English mechanics. The resulting data were used to determine the reliability of the essay grading procedures employed in connection with the Alberta departmental examinations (Nyberg, 1966). From the sample of 103 essays, models representing stanines 2, 4, 6, and 8 were chosen for style and content, and another set also representing stanines 2, 4, 6 and 8 was selected for use in grading mechanics of English. Reasons for developing two scales instead of one, and details of how the scales were developed are provided in the technical report (Nyberg & Nyberg, 1978).

When the two scales, style-content and mechanics, consisting of model essays, had been completed, the question remained as to whether they could be used by teachers to grade students' essays with at least some degree of reliability. The study reported here was designed to answer this question.

Procedure

The general procedure was quite simple. A set of essays, written by English 30 students, was rated independently by each of six teachers through use of the models. The mean correlation between all pairs of raters was then calculated.

In April of 1977, four teachers administered an essay test to their English 30 classes. The test consisted of writing a 200-word essay on the topic "The Ideal School." A time limit of one hour was imposed. The students were from unselected classes; that is, they ranged in achievement and ability from low to high. In order to

motivate the students to perform at their best, it was announced that the test results would contribute to their final course grades. Eighty-three essays were turned in for use in this study.

Six experienced teachers of high school English each graded the full set of essays. Each of the teachers was provided with a copy of the models to be used for rating and each was asked to practise using the models as suggested in the scale booklet. Each rater then graded each essay twice — once for style-content, and once for mechanics. No marks were placed on the students' test papers; instead, scores were entered on grading sheets retained by the raters. This procedure prevented a reader from knowing how any essay fared at the hands of the other readers. Each essay was identified by number only; therefore it was impossible for the rater to know who had written the essay or from which school it came.

The correlation between scores awarded by each pair of raters was determined. The mean of all the correlations between pairs of raters was taken as an estimate of the reliability of the grading procedures.

The means and standard deviations of the scores awarded by the six raters over the 83 essays were calculated and then examined for differences.

Results

Table 1 shows the correlations between all pairs of raters for English mechanics and Table 2 shows the means and standard deviations. Tables 3 and 4 provide similar information for the style-content grades.

The reliability of the scoring for mechanics, estimated by computing the mean of the correlations in Table 1, was found to be .782. For style-content, a reliability estimate of .772 was calculated. These values compare favourably with estimates from other studies. Standard errors of measurement were estimated to be .659 for style-content and .751 for mechanics. In the study from which the essay models were derived, Nyberg (1966) reported reliabilities of .768 for mechanics and .595 for style-content. It was concluded that a respectable reliability of scoring can be achieved through use of the essay models by teachers rating twelfth-grade essays.

Analysis of variance indicated no significant differences ($p < .05$) among the raters when they graded for mechanics, but for the style-content component, the differences were highly significant. This meant that some of the raters consistently graded essays higher (or lower) than other raters on the style-content aspect. This is of no particular consequence to teachers dealing only with their own classes;

TABLE 1

CORRELATIONS BETWEEN MECHANICS SCORES AWARDED BY SIX RATERS

	Rater Number				
Rater Number	2	3	4	5	6
1	.949	.836	.743	.880	.830
2		.823	.672	.821	.809
3			.592	.738	.689
4				.837	.698
5					.820

TABLE 2
MEANS AND STANDARD DEVIATIONS OF MECHANICS SCORES

Rater Number	Means	Standard Deviation
1	5.096	1.588
2	4.964	1.594
3	5.193	1.548
4	5.012	1.654
5	5.048	1.575
6	5.277	1.688

TABLE 3
CORRELATIONS BETWEEN STYLE-CONTENT SCORES AWARDED BY SIX RATERS

	Rater Number				
Rater Number	2	3	4	5	6
1	.952	.735	.768	.903	.844
2		.704	.757	.866	.815
3			.624	.671	.609
4				.822	.687
5					.820

TABLE 4
MEANS AND STANDARD DEVIATIONS OF STYLE-CONTENT SCORES

Rater Number	Means	Standard Deviation
1	5.759	1.359
2	5.699	1.462
3	6.470	1.320
4	5.181	1.300
5	5.470	1.329
6	5.940	1.563

however, in cases where the models are being used to grade essays for purposes of comparison where several raters are involved, measures would have to be taken to reduce or eliminate the differences. This might be done through discussions among the teachers to resolve the differences between the "hard" and the "soft" raters. In this study, there was no group discussion of the grading. A simple mathematical correction could be employed to eliminate differences among raters. Converting all ratings to T scores is one of the procedures that would serve the purpose.

Concluding Statement

The reliability of scoring through use of the essay models was quite satisfactory in view of results obtained in other studies. Nyberg (1978, p. 8), after reviewing the literature, stated: "Reliability coefficients involving senior high school students and several classroom teachers as raters are seldom higher than .6." This should not be taken to mean that the reliability coefficients were high. Values of .782 and .772 would be considered weak in mathematics problem-solving tests or in multiple choice tests in, say, science. It must be emphasized, also, that this study investigated reliability of SCORING only, which must not be confused with TEST reliability. That is to say, the only source of unreliability that was studied here was within the readers, while test reliability includes variation from one situation to another.

The primary value of the Alberta Essay Scales would be to assist the classroom teacher in day-to-day scoring of essays by providing a constant standard. The scoring procedures are not sufficiently reliable to permit decisions regarding individual students to be made on the basis of one essay. However, if, as is ordinarily the case, final grades are based on at least a dozen writing assignments, decisions for individuals would be reasonably sound.

Verner R. Nyberg
Department of Educational Psychology
The University of Alberta

and

Adell M. Nyberg
Spruce Grove Composite High School

Copies of the Alberta Essay Scales are available from the Curriculum Branch, Alberta Education.

References

Buros, O. K. *Sixth Mental Measurement Yearbook.* Highland Park, N.J.: Gryphon Press, 1965.

Educational Testing Service. *Cooperative Sequential Tests of Educational Progress Handbook for Essay Tests Level 2.* Princeton, N.J.: Educational Testing Service, 1957.

Hillegas, M. B. *A scale for the measurement of quality in English for young people.* New York: Bureau of Publications, Teachers' College, Columbia University, 1912.

Hudelson, E. *Hudelson English Composition Scale.* New York: World Book Company, 1923.

Nyberg, A. M. *A study of the reliability of an Alberta composition scale.* Unpublished M.Ed. Colloquium Report, University of Alberta, 1977.

Nyberg, V. R. *A factor analytic study of essay gradings.* Unpublished Ed.D. dissertation, U.C.L.A., 1966.

Nyberg, V. R., & Nyberg, A. M. *Alberta Essay Scales.* Edmonton, Alberta: Alberta Education, 1977.

Nyberg, V. R., & Nyberg, A. M. *Technical report on Alberta Essay Scales.* Edmonton, Alberta: Alberta Education, 1978.

Van Wagenen, M. J. *Van Wagenen English Composition Scales.* New York: World Book Company, 1923.

Willing, M. H. *Willing Scale for Measuring Written Composition.* Bloomington, Ill.: Public School Publishing Co., 1922.

BOOK REVIEWS

THE REFORM OF TEACHER EDUCATION IN THE UNITED KINGDOM. *By James Lynch.* Guildford, Surrey: Society for Research into Higher Education, 1979, 211 pp. (ISBN 0-900868-68-6)

This is essentially a book about the role played in teacher education by the Council for National Academic Awards, and about the advantages of "a modular course system" for teacher training programmes. On these topics, the book is very well documented and provides, in text and appendices, detailed information that would be of use to course planners.

The Council for National Academic Awards is a body set up by the Department of Education and Science for the express purpose of validating courses offered by institutions of higher education not entitled to validate their own, that is to say, non-university institutions. It is the keystone of the structure built up in the U.K. over the past twenty-five years to provide higher education outside the universities. Together the two systems constitute "the binary system" of higher education, and the key to the difference lies in the fact that universities in this country are funded by the University Grants Committee which draws its resources directly from the Treasury, whereas polytechnics and other non-university institutions are funded through channels negotiated by the Department of Education and Science with the local education authorities. The split began on a substantive basis, responsibility for the development of higher technical education being removed from the domain of the universities, which continued to be responsible for the development of academic studies. That difference has since been steadily eroded, as is indicated by the fact that the original National Council for Technological Awards, set up in 1955, became in 1964 the National Council for Academic Awards. While the author of this book recognizes that central government has been involved in a "power game" (p. 2), he seems to see no threat in the take-over of teacher education under cover of the developing binary system. He indicates instead that government control of manpower in the education system necessitated the move.

A reader of Chapter 1 (the historical context) and Chapter 2 (contemporary developments) might be forgiven if he nurtured a growing sense that the reference to "reform" in the author's title constituted a grand piece of irony. Predictions of sharply falling school enrolments are, of course, a major factor in the situation described, but there are many other sources of uncertainty and confusion, and the outcomes are vacillation, delay, panic action, administrative muddle, and a generally low morale.

But with Chapter 3 (the expansion of C.N.A.A. activities), there is a marked change of tone: it is clear that the Fairy Prince has now entered upon the scene and

the stage is set for the promised reform. The Fairy prince is, of course, the Council for National Academic Awards, and it seems that the magic power by which he will deliver the system from bondage lies principally in provision for modular courses. Of the C.N.A.A. we are told:

> It has been argued that, in terms of its democratizing and academic impact, the Council represented an innovatory development which far surpassed the contribution of the new universities founded in the 1960s and even the much-vaunted Open University, and which evolved without the massive injections of capital that those institutions required. Moreover, its foundation marked the beginning of an important development in teacher education towards the sharing of responsibility, for it gradually weaned much of teacher education from the autonomous to the public sector of higher education, from an "élitist" towards a "mass" tradition of further and higher education. (p. 58)

Chapter 4 spells out the curricular implications of this "innovatory development" and Chapter 5, its organizational implications. Chapter 6 is a brief look at the shape of things to come, and here an odd kind of Utopia comes into view:

> Increased control directly from central government and indirectly through the regional councils . . . will enable the central government to regulate not only teacher education numbers and courses but also other courses in higher education, and while it is not likely to stem the greater variety of course provision which has been stimulated by the entry of C.N.A.A. to the validation field, it will nonetheless have effects on the curricula of the institutions of higher education and also on the curricula of the schools; effects which remain as yet relatively unappreciated. Such developments carry important, if indirect, implications for the control of school curricula by central government, and thus also for the distribution of knowledge in society. (p. 140)

Let me make it clear that I share the author's respect for what the C.N.A.A. has achieved, and that I have for many years been concerned, as he is, to develop innovatory courses which would introduce a stronger element of professionalism into teacher education. That said, the biased nature of his account must be recognized. What the universities have contributed, both through their education departments and more generally, and what colleges of education have achieved and continue to achieve in cooperation with the universities is almost totally ignored. The "waning influence" of universities is greeted as an aspect of reform, and the reader is led to believe that, thankfully, the university phase in teacher education is over and we need think no more about it. A writer who can equate university education with "élitism" and take-over by government as "democratization," and who is disposed to conflate manpower planning with curricular control, is surely taking a myopic view of the role of education in society. Such a view of teacher education would have been challenged had he raised his eyes even so far as to glance at a neighbouring profession, that of the medical service.

This book was originally written in German and published in Germany: it is to be regretted that so partial a view of teacher education in the United Kingdom should be exported, whether to Germany or to the American continent.

James Britton
Emeritus Professor of Education
in the University of London

TEACHING CANADA FOR THE '80s. *By A. B. Hodgetts and Paul Gallagher.* Toronto: The Ontario Institute for Studies in Education, 1978, 136 pp.

Hodgetts and Gallagher indicate that their book is "presented to educators and all other interested Canadians in every part of the country as a guide — or at least a starting point — to teaching about Canada in the next decade" (p. 6). This guide is comprised of seven chapters in which the authors attempt to define the nature of Canada studies, outline possible content of Canada studies programs, and identify the unique roles of the elementary and secondary schools in achieving the goals of such programs. In addition, chapter seven sets forth a particular framework for approaching instruction of public issues in secondary school classrooms. How useful the guide will be to educators with concerns in these areas is open to question.

In attempting to arrive at a sound conceptual base for their scheme, the authors have run into definitional problems. The impression is left that there has been no real attempt to explore the idea of Canadian or Canada studies but that the authors armed with a preset notion of what content students should acquire have dispensed rather quickly and inadequately with the rationale for such studies. An example of the resultant confusion may be seen in conflicting statements such as:

> We believe that *Canada* studies — as distinct from all other more restricted *Canadian* studies should deal with our society as a totality, in country-wide, interlocking perspectives. (p. 2)

followed only a few pages later by:

> A clear distinction may now be made between Canadian studies and Canada studies. The adjective "Canadian" described a much broader field of study, including all kinds of local or specialized concerns, as well as a wide range of human activities of no direct political significance. (p. 11)

Further confusion is created when Canada studies are next equated with civic education which is "designed to encourage a development of the skilled and sensitive public opinion needed to resolve deep-seated differences in the Canadian political community" (p. 11).

Semantics aside, the proposed Canada studies program seems limited in scope and richness. The social science perspective put forward lacks the human concern so evident in proposals such as that of Symons (1975). For Symons, it is important that Canadians develop self-knowledge, an understanding of themselves, who they are, where they are in time and space, where they have been, where they are going, what they possess, and what their responsibilities are to others. There is as much concern for Canadian art, literature, and philosophy as there is for the functioning of parliament and the mechanics of the economic system. Symons' proposals stir the imagination and challenge the intellect. In contrast, the text-like proposals set forth in the book, which begin in the elementary school with a study of basic citizenship and Canadian communities, move through a concern with the "total environment of Canada" at the beginning of secondary school to the Canadian political and economic systems, and finally culminate with the examination of public issues in the last years of the secondary school, seem pedantic and uninspiring.

The limitations of the range and scope of Canada studies proposed are coupled with an equally inadequate approach to knowledge. Despite notable contradictions within the text, the dominant view of knowledge is that described by Freire (1971) as the "banking concept" in which the teacher's task is to deposit with the student what is considered to be the truth about Canada. Faith in this approach is characterized by statements such as:

> Yet it is primarily by raising the knowledge levels of our students that we may hope to achieve the kinds of understandings and respect for others that are the attributes of a civilized people. (p. 3)

Pages 38 through 117 of the book follow this approach by outlining theories, themes, and generalization which, the authors indicate, students should learn about Canada. There are close to 200 statements predicated with the stem "Students should understand" or some equivalent variant indicating a specific piece of knowledge for which students should be held accountable.

This view of knowledge, popularized by Bruner and so characteristic of the social studies projects of the early sixties, is bound to disappoint those who view such approaches to knowledge building as questionable. The knowledge base proposals not only seem to lack the support of many current theorists, but also to ignore the results of important contemporary studies in the area of civic education such as that of Torney, Oppenheim, and Farnen (1975) who have indicated that:

> The acquisition of knowledge does not correlate highly with support for democratic values. . . . Neither does knowledge of civics correlate highly with various measures of political participation. (p. 19)

Thus, even if youngsters were to acquire the knowledge base outlined, the authors' stated purpose of encouraging "Education for citizenship — for an active role in the life of society" (p. 19) would not seem to be assured.

There is a sharp contrast between the suggested programs for the elementary grades and that proposed for the secondary schools in the amount of specification provided by the authors. The elementary program outlined begins with primary grade children exploring areas such as community participation, the basic responsibilities of community members, and an examination of public controversy. Out of these studies, it is suggested, children will develop attitudes toward Canada rather than accumulate information. It is further argued that the present content of social studies programs in the primary grades is suitable — all that is required to turn it into a Canada studies program is a new emphasis on the local community as a dynamic social reality.

In the middle and upper years of the elementary school, children will study community life in Canada and the world, Canada as a political community, Canada in the world community, participative citizenship, compromise and consensus, and representative government. There is also a plea for history and geography of Canada in the last years of the elementary school, but how this fits into the total scheme is left unclear. It is expected that from such a program children will emerge from the elementary school with "a real sense of personal identification with Canada as a whole," and will be ready for "a more systematic study of the basic components of Canada studies" (p. 33).

The junior high school is to be primarily devoted to a study of the Canadian Environment, focussing on the development of five major generalizations:

1. Canada is a northern, vast, and regionally divided country.
2. Canada has a broad natural resource base composed of both renewable and non-renewable resources.
3. Canada is an industrial, technological, and urbanized society.
4. Canada is a culturally diverse, multi-ethnic country with two historically dominant linguistic and cultural groups.
5. Canada is exposed to a multitude of external economic, political, and cultural influences. (p. 12)

By the end of the junior high school years, it is contended, students will have an understanding of the complexity of Canada and an awareness of the responsibilities of citizenship in operating the Canadian political community.

Students in the ninth to twelfth grades, approximately, are to build upon this base knowledge of *what* Canada is, by moving to *how* Canada functions. This is to be accomplished through the examination of the sytems that Canadians have developed, primarily the economic and political systems. Following this detailed study of systems, students move to a study of public issues as the climax to all of their preceding work on Canada studies.

The rationale and suggestions for the elementary school program are dealt with in 24 pages, the early secondary school in 16 pages, while the middle and upper secondary programs are detailed in 75 pages. Those concerned with elementary education will probably find the suggestions interesting but superficial; those interested in secondary education will probably find the detailed suggestions highly restrictive and overprescriptive. All will likely find that the proposals provide a limited explanatory framework of Canada and Canadian development. This is due in part to the interpretation of Canada within particular political and economic models and the extent to which human experience elsewhere is ignored.

The proposed framework raises serious concerns about the nature of Canada studies and the imbalances that could occur in youngsters' social learnings were such a plan to be followed in schools. The authors are aware that their scheme may be confining, at least in terms of geographic settings explored:

> It is legitimate to contend that an exclusive preoccupation with Canada and things Canadian throughout the elementary years is too narrow and chauvinistic. (p. 28)

However, they do not seem to be sensitive to the possibility that the approach taken to knowledge and knowledge building may be equally restrictive.

Teachers will find the model set forth in the last chapter the part of the book most readily put into practice. However, before doing so, they will want to consider carefully the inquiry model proposed. It consists of a series of six questions set forth under three headings, Contemporary Analysis, Historical Analysis, and Future Analysis. It is suggested that this is "a total package for studying any major public issue in Canada" (p. 123).

The assumptions underlying the model have not been set forth by the authors. In questioning the effect that extended use of such a model might have on students, one would hopefully have considered that it seems to support the following: (1) Human problems can be reduced and managed with the application of a particular inquiry model, (2) If only the "correct" questions are asked then issue resolution will be assured, (3) Students are incapable of exploring issues without the aid of a recipe, (4) One model for inquiry into issues is satisfactory, (5) The scientific-technical model proposed is appropriate for all issues, (6) Application of the model absolves students of any further intellectual responsibility, (7) Issue exploration for students does not proceed past the appraisive level. Many will find these underlying assumptions worthy of examination before proceeding to use the model in classrooms.

One of the other obvious shortcomings of the model is that it makes no attempt to alert students to the methodology-in-use. The model lacks what Goldmark (1968) has labelled as inquiry-into-inquiry. This essential component of knowledge building asks that learners explore what they did, how they did it, why they did it that way, and if they should do it that way in the process of problem-solving. This inquiry into the methodology used to explore issues would seem necessary if students are to do other than to apply unwittingly, in a mechanistic manner, questions which will eventually influence what they are able to think and the very way in which they will ultimately structure their own reality.

Advocation of Canadian content in school programs boasts wide and vocal sup-

port. It represents a concern over the lack of a national consciousness and a struggle for a distinctive Canadian identity free from colonial or neo-colonial influence. *Teaching Canada for the 80's* represents one of the few attempts to describe a comprehensive program for schools in this area. It would be unfortunate if, in our zest to be Canadian, the obligation to be scholarly were compromised.

D. Massey
Department of Elementary Education
The University of Alberta

References

Freire, P. *Pedagogy of the oppressed.* New York: Harder and Harder, 1971.

Goldmark, B. *Social Studies a method of inquiry.* Belmont: Wadsworth, 1968.

Torney, J. V., Oppenheim, A. N., & Farnen, R. F. *Civic education in ten countries.* New York: John Wiley and Sons, 1975.

Symons, T. H. *To know ourselves.* Report of the Commission on Canadian Studies. Ottawa: Association of Universities and Colleges, 1975.

EDUCATIONAL PSYCHOLOGY. *By David W. Johnson.* Englewood Cliffs: Prentice Hall, 1979, 592 pp., $12.95 (U.S.).

The three main subdivisions of this omnibus introductory text in educational psychology are before teaching, during teaching, and after teaching. Four chapters each are devoted to before and after teaching, while 9 chapters outline the psychological concepts of importance during teaching. The before teaching section sets out the scope of educational psychology, the crisis in educational goals, cognitive and intellectual development, and socialization and social development. Major theorists and their theories are set out in sufficient detail for beginning students. The 9 chapter section dealing with during teaching outlines in rather pragmatic form the problems involved in specifying learning goals and outlines both the cognitive and the behavioural approaches to instruction. Johnson also deals with attitude change, motivation, and learning environments to ensure skills and abilities are available to users of his text. He specifically deals with conflicts and classroom control and discipline, topics that are often sparsely treated in many educational psychology texts. A section dealing with students who have special needs, while not too appropriate for special education students, is quite adequate for the classroom teacher. The after teaching section deals with problems in evaluating students, evaluating instructional effectiveness, a sociologically oriented chapter dealing with improving your school and school district, and finally a synoptic statement dealing with your career as a teacher.

Johnson states that the book does not assume any prior knowledge of psychology, that the approach is eclectic with differing points of view presented and no single stance being adopted. There appears, however, to be a social-psychological stance that runs throughout the entire book. This should not be surprising with Johnson's interest in the areas of social psychology and interpersonal relations. While there is a claim that the units are self-contained, there are circumstances that require and/or direct the reader back to earlier sections of the book. The group cooperative activities approach to instruction and learning — one that is endorsed within the text itself — is obvious from the many student activities that are presented. This represents a serious attempt to bridge the gap between theory and practice and is accomplished quite successfully. No

student workbook is needed which is a positive feature. An *Instructor's Manual* is apparently available, but it was not received for this review. Johnson's approach is based upon 14 years of teaching experience and other professional work that he has been associated with. Other positive points include the highlighting of key concepts during printing, and the inclusion of incidental information relating to several topics under discussion that is not critical to the textual material, but certainly does expand upon the concept for those interested. Group activities spaced throughout the book require the students to clarify their own concepts and to reconcile their concepts with the theoretical statements that they are studying. Chapter summaries are also included.

Despite many of the claims by the author, there are some obvious attempts to present a particular point of view. For example, Gagné's taxonomy is the only usable structure for stating educational objectives. Bloom is given passing mention, but no attempt has been made to place his work in historical perspective. Missing also is the listing of the *Taxonomy of Educational Objectives: Handbook II,* and Harrow's 1972 work on *A Taxonomy of the Psychomotor Domain.*

The after teaching section of the book dealing with evaluating students should be improved. Measurement and evaluation are characterized as unpleasant tasks due to negative student reaction, yet properly handled they can be an asset to any teacher and will be appreciated by the students. Without evaluation, students would receive no feedback and would be unable to assess their relative position within the school setting — something they and their parents insist upon doing. The section dealing with statistics is far too short — a separate appendix would be more appropriate — and contains an incorrect definition of the standard deviation. Technical qualities of instruments are stated in order of importance — reliability, validity, objectivity, practicality and discrimination. Many writers would instead list validity, reliability and usability.

Standardized tests are covered in some detail, but the standard reference work, O. K. Buros and his seven editions of the *Mental Measurements Yearbook,* are not mentioned, nor are they included in the references. This omission is a serious error as without knowledge of the Buros publication, choosing a standardized test efficiently is virtually impossible. A brief section on teacher-made tests includes suggestions for item construction. The matching item on p. 468, however, ignores this advice.

Since 1968, much attention has been devoted to the emerging concepts involved in educational evaluation as exemplified in the more recent publications of R. Stake, D. Stufflebeam, M. Scriven, and M. Provus. This entire field of study has simply been omitted and as it constitutes a significant area of study today, it should be included in an introductory text. Many improvements that have occurred in the school system can be attributed to evaluation studies and this makes the area too important to ignore.

On balance, the first thirteen chapters (before and during teaching sections) are for the most part quite appropriate for an introductory course. There is a glossary which includes most of the psychological terms the student encounters, as well as a References section that lists over 750 books and articles. For most beginning students this unselected reference list will be somewhat formidable, but many valuable publications are listed therein. There is also a workable index. Johnson writes in a very readable fashion, and if you subscribe to his social-psychological approach, you will find much of value for your students.

Frederick C. Rankine
Professor of Education
University of New Brunswick

PRINTED BY
THE UNIVERSITY OF ALBERTA
PRINTING DEPARTMENT